(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number
WO 02/24736 A1

(51) International Patent Classification[7]: C07K 14/195, C07H 21/04, C12N 15/52, 15/62

(21) International Application Number: PCT/AU01/01190

(22) International Filing Date: 21 September 2001 (21.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

PR 0277	21 September 2000 (21.09.2000)	AU
PR 0304	22 September 2000 (22.09.2000)	AU
PR 0320	22 September 2000 (22.09.2000)	AU

(71) Applicant *(for all designated States except US)*: **THE UNIVERSITY OF QUEENSLAND** [AU/AU]; St Lucia, Queensland 4067 (AU).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **BIRCH, Robert** [AU/AU]; 24 Yarimbah Street, Jindalee, Queensland 4074 (AU).

(74) Agents: **ARGAET, Victor, Peter** et al.; Davies Collison Cave, Level 3, 303 Coronation Drive, Milton, Queensland 4064 (AU).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH ANTIBIOTIC BIOSYNTHESIS AND USES THEREFOR

(57) Abstract: The present invention discloses polyketides and the polyketide synthases and ancillary enzymes that are capable of producing such compounds. More particularly, the present invention discloses polynucleotides and polypeptides associated with (i) a novel polyketide synthase linked to a non-ribosomal peptide synthetase involved in the biosynthesis of albicidins, (ii) a novel phosphopantetheinyl transferase for activating enzymes, particularly polyketide synthases and/or non-ribosomal peptide synthetases, associated with the biosynthesis of albicidins, and (iii) a novel methyltransferase for methylating precursors of albicidins and/or intermediates related to albicidin biosynthesis. The present invention also discloses methods of using the aforementioned polynucleotides and polypeptides for activating polyketide synthases and/or non-ribosomal peptide synthetases, for methylating precursors of albicidins or their analogues and/or intermediates involved in the biosynthesis of albicidins or analogues thereof and for enhancing the level and/or functional activity of albicidins or their analogues. Also disclosed are methods of using the polynucleotides and polypeptides of the invention for the biosynthesis of albicidins or their analogues.

WO 02/24736 A1

POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH ANTIBIOTIC BIOSYNTHESIS AND USES THEREFOR

FIELD OF THE INVENTION

THIS INVENTION relates generally to antibiotic biosynthesis. More particularly, the present invention relates to polyketides and the polyketide synthases and ancillary enzymes that are capable of producing such compounds. Even more particularly, the present invention relates to a polyketide synthase linked to a non-ribosomal peptide synthetase involved in the biosynthesis of albicidins, to a phosphopantetheinyl transferase for activating enzymes, particularly polyketide synthases and/or non-ribosomal peptide synthetases, associated with the biosynthesis of albicidins, and to a methyltransferase for methylating precursors of albicidins and/or intermediates related to albicidin biosynthesis. The present invention also relates to biologically active fragments of the aforementioned polypeptides and to variants and derivatives of these molecules. Further, the invention relates to polynucleotides encoding the said polypeptides, including the *xabA*, *xabB* and *xabC* genes of *Xanthomonas albilineans*, to polynucleotides encoding the said fragments, variants or derivatives, to vectors comprising the said polynucleotides and to host cells containing such vectors. The invention also relates to a transcriptional control element for modulating the expression of polynucleotides including, for example, the *xabB* gene and/or the *xabC* gene of *Xanthomonas albilineans*, or variants thereof. The invention also features methods of using the polynucleotides, polypeptides, fragments, variants, derivatives and vectors for activating polyketide synthases and/or non-ribosomal peptide synthetases, for methylating precursors of albicidins or their analogues and/or intermediates involved in the biosynthesis of albicidins or their analogues and for enhancing the level and/or functional activity of albicidins or their analogues. The invention also encompasses methods of using the aforesaid polynucleotides, polypeptides, fragments, variants and derivatives for the biosynthesis of albicidins or analogues thereof.

Bibliographic details of various publications referred to by author in this specification are collected at the end of the description.

BACKGROUND OF THE INVENTION

Polyketides represent a large structurally diverse group of compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. They possess a broad range of biological activities including antibiotic and pharmacological properties. For example, polyketides are represented by antibiotics such as tetracyclines, erythromycins, immunosuppressants such as FK506, FK520 and rapamycin, anticancer agents such as daunomycin and veterinary products such as monensin and avermectin.

Considering the difficulty in producing polyketide compounds by conventional chemical methodologies, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds. In this regard, reference may be made to PCT publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; WO 97/02358; and WO 98/27203; U.S. Pat. Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; and 5,712,146; Fu *et al.* (1994, *Biochemistry* **33**: 9321-9326); McDaniel *et al.* (1993, *Science* **262**: 1546-1550); and Rohr (1995, *Angew. Chem. Int. Ed. Engl.* **34**(8): 881-888).

Polyketides are synthesised in nature by polyketide synthases (PKS). These enzymes, which are actually complexes of multiple enzyme activities, are in some ways similar to, but in other ways different from, the synthases that catalyse condensation of 2-carbon units in the biosynthesis of fatty acids. Specifically, PKS enzymes catalyse the biosynthesis of polyketides through repeated (decarboxylative) Claisen condensations between acylthioesters (*e.g.*, acetyl, propionyl, malonyl or methylmalonyl). Following each condensation, they introduce structural variability into the product by catalysing all, part, or none of a reductive cycle comprising a ketoreduction, dehydration, and enoylreduction on the β-keto group of the growing polyketide chain. PKS enzymes incorporate enormous structural diversity into their products, in addition to varying the condensation cycle, by controlling choice of primer, extender units, and the overall chain length and, particularly in the case of aromatic polyketides, regiospecific cyclisation of the nascent polyketide chain. After the carbon chain has grown to a length characteristic of each specific product, it is released from the synthase by thiolysis or acyltransfer. Thus, the PKS complexes consist of families of enzymes which work together to produce a given polyketide. It is the choice of chain-building units, controlled variation in chain length, and the reductive cycle,

genetically programmed into each PKS, that contributes to the variation seen among naturally occurring polyketides.

Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesised. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes. These classifications are well known and reference may be made, for example, to Hopwood and Khosla (1992).

The Type I or modular PKS enzymes typically catalyse the biosynthesis of complex polyketides such as erythromycin and avermectin. These modular enzymes include assemblies of several large multifunctional proteins carrying, between them, a set of separate active sites for each step of carbon chain assembly and modification (Cortes *et al.*, 1990; Donadio *et al.*, 1991; MacNeil *et al.*, 1992). Accordingly, modular PKS complexes can be viewed as biochemical assembly lines, composed of a series of catalytic domains involved in sequential assembly and modification of acyl groups on the growing polyketide chain (Cane *et al.*, 1998; Keating and Walsh, 1999). The catalytic domains are arranged in "modules", punctuated by acyl carrier protein (ACP) domains that tether the nascent polyketide while it undergoes the catalytic modifications programmed in the associated module. For each polyketide there is an initiation module, a series of elongation modules that define the length and structure of the polyketide chain, and a termination module to release the product from the final tether. The initiation module typically comprises an acyl transferase (AT) domain that couples the initial acyl group from an acyl-CoA substrate to the phosphopantetheinyl tether of the first ACP domain. Each elongation module typically comprises a ketosynthase (KS), an AT and an ACP. The KS removes the growing polyketide unit from the upstream ACP and couples it to the next acyl group in the chain, which has already been selected and loaded by the AT onto the ACP in the same module. Other catalytic domains (*eg.* a ketoacyl reductase (KR), and dehydratase (DH)) within an elongation module can modify the newly elongated polyketide before it is transferred to the next module in the biochemical assembly line. A thioesterase (TE) domain in the termination module accomplishes release of the assembled polyketide from the last ACP in the series (Cane *et al.*, 1998; Keating and Walsh, 1999).

Biosynthesis of a polyketide can involve the sequential action of several PKS proteins, each with one to six elongation modules (MacNeil *et al.*, 1992; Apricio *et al.*, 1996). There are variations on the modular PKS design, including participation by some loading domains across modules or in *trans* from separate proteins (Keating and Walsh, 1999), and several examples of hybrid PKS/NRPS proteins (Albertini *et al.*, 1995; Gehring *et al.*, 1998; Duitman *et al.*, 1999; Paitan *et al.*, 1999). Subsequent modification of the polyketide by dedicated tailoring enzymes is generally required to complete the biologically active product (Hopwood, 1997). Other biologically active compounds including antibiotics comprise polypeptides assembled by non-ribosomal peptide synthetases (NRPSs). NRPSs typically show a modular architecture and tethered biosynthetic strategy analogous to PKSs (Cane *et al.*, 1998; Keating and Walsh, 1999). In NRPSs a condensation (C) domain removes the growing peptide unit from the upstream PCP domain and couples it to the next amino acid group in the chain, which has already been selected and loaded by an adenylation (A) domain onto the PCP in the same module (Marahiel *et al.*, 1997; Stachelhaus *et al.*, 1998). Other catalytic domains (*e.g.*, epimerase or N-methytransferase) within an elongation module can modify the newly elongated polypeptide before it is transferred to the next module in the biochemical assembly line (Marahiel *et al.*, 1997).

Many phytopathogenic bacteria and fungi secrete toxins with phytotoxic activity and a broad spectrum of antimicrobial properties (Guenzi *et al.*, 1998). Albicidin phytotoxins are polyketides produced by *Xanthomonas albilineans*, which are key pathogenicity factors in the development of leaf scald, one of the most devastating diseases of sugarcane (*Saccharum*, interspecific hybrids) (Ricaud and Ryan, 1989; Zhang and Birch, 1997; Zhang *et al.*, 1999). Albicidins selectively block prokaryote DNA replication and cause the characteristic chlorotic symptoms of leaf scald disease by blocking chloroplast development (Birch and Patil, 1983; 1985b; 1987a; 1987b). Because albicidins are rapidly bactericidal at nanomolar concentrations against a broad range of Gram-positive and Gram-negative bacteria, they are also of interest as potential clinical antibiotics (Birch and Patil, 1985a).

The major antimicrobial component of the family of albicidins produced in culture by *X. albilineans* has been partially characterised as a low M_r compound with several aromatic rings (Birch and Patil, 1985a). Low yields have slowed studies into the

chemical structure of albicidin, its application as a tool to study prokaryote DNA replication, and its development as a clinical antibiotic (Zhang *et al.*, 1998). Genetic analysis of albicidin biosynthesis is likely to indicate approaches to increase yields, probable structural features, and opportunities for engineering novel antibiotics in this family.

SUMMARY OF THE INVENTION

The present invention arises in part from the identification and characterisation of several *X. albilineans* genes associated with albicidin biosynthesis. In particular, the present inventor has isolated a novel *X. albilineans* gene (*xabB*), which encodes a large protein (predicted *M*r 525,695), with a modular architecture indicative of a multifunctional PKS linked to a non-ribosomal peptide synthetase (NRPS). At 4801 amino acids in length, the product of *xabB* (XabB) is the largest reported PKS-NRPS. Twelve catalytic domains in this multifunctional enzyme are arranged in the order N-terminus-acyl-CoA ligase (AL)-acyl carrier protein (ACP)-β-ketoacyl synthase (KS)-β-ketoacyl reductase (KR)-ACP-ACP-KS-peptidyl carrier protein (PCP)-condensation domain (C)-adenylation domain (A)-PCP-C. The modular architecture of XabB indicates likely steps in albicidin biosynthesis, and approaches to enhance antibiotic yield. The novel pattern of domains, in comparison with known PKS-NRPS enzymes for antibiotic production, also contributes to the knowledge base for rational design of enzymes producing novel antibiotics. The present inventor has found that XabB is required for the production of albicidins and that enhanced expression of *xabB* leads to increased levels and/or functional activities of albicidin antibiotics.

A gene (*xabC*) encoding a novel *O*-methyltransferase has also been isolated, which methylates albicidin precursors and/or intermediates involved in albicidin biosynthesis. Surprisingly, enhanced expression of *xabC* has been found to increase the levels and/or functional activities of albicidin antibiotics.

The present inventor has also isolated a gene (*xabA*) encoding a phosphopantetheinyl transferase (PPTase), which is required for post-translational activation of synthetases in the albicidin biosynthetic pathway. In this regard, it is known that inefficient phosphopantetheinylation has limited the activity of other antibiotic synthetases overexpressed in heterologous species (Walsh *et al.*, 1997). Accordingly, the isolated *xabA* gene, together with its target in the albicidin biosynthetic pathway (*e.g.*, *xabB*), provide the means to engineer high level co-expression of the albicidin synthetase and its activating PPTase to obtain albicidins in higher yields, and ultimately to manipulate the elements of the albicidin biosynthetic machinery, by mutagenesis or by other means, to produce desired structural variants of this novel antibiotic class.

The above genes, in whole or in part, together with their variants and derivatives, are useful *inter alia* for modulating the level and/or functional activity of albicidins, for expressing PKS enzymes in recombinant host cells, for producing polyketides including albicidins and their analogues and for combinatorial biosynthesis, as described hereinafter.

Accordingly, one aspect of the present invention contemplates an isolated polypeptide encoding at least a portion of an albicidin PKS-NRPS (XabB) or its variants or derivatives. In one embodiment of this type, the invention provides an isolated polypeptide comprising at least one domain selected from the group consisting of:

(a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.

(b) a β-ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;

(c) a β-ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;

(d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;

(e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof;

(f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, or variants thereof; and

(g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

Preferably, the AL domain comprises each of the sequences set forth in SEQ ID NO: 6 and 8, or variants thereof.

In one embodiment, the KS domain preferably comprises each of the sequences set forth in SEQ ID NO: 10, 12 and 14, or variants thereof. In an alternate embodiment, the KS domain preferably comprises each of the sequences set forth in SEQ ID NO: 16, 18 and 20, or variants thereof.

Preferably, the A domain comprises each of the sequences set forth in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

In one embodiment, the C domain preferably comprises each of the sequences set forth in SEQ ID NO: 54, 56, 58, 60, 62, 64 and 66, or variants thereof. In an alternate embodiment, the C domain preferably comprises each of the sequences set forth in SEQ ID NO: 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

In another embodiment, the invention provides an isolated polypeptide comprising at least a biologically active fragment or portion of the sequence set forth in SEQ ID NO: 2, or a variant or derivative thereof.

Suitably, the biologically active fragment is at least 6 amino acids in length.

In a preferred embodiment, the domains broadly described above are arranged in an N- to C-terminal direction as follows: AL-ACP-KS-KR-ACP-ACP-KS-PCP-C-A-PCP-C.

Suitably, the biologically active fragment comprises at least one domain selected from the group consisting of the AL domain, the KS domain, the KR domain, the ACP domain, the A domain, the PCP domain and the C domain as broadly described above.

Suitably, the variant has at least 60%, preferably at least 70%, more preferably at least 80%, more preferably at least 90% and still more preferably at least 95% sequence identity to the sequence set forth in SEQ ID NO: 2.

Preferably, the variant comprises at least one sequence selected from the group consisting of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variant thereof. In this regard, the variant preferably has at least 70%, preferably at least 80%, more preferably at least 90%, and still more preferably at least 95% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80.

In another aspect, the present invention contemplates an isolated polypeptide encoding at least a portion of a PPTase (XabA) associated with albicidin biosynthesis or its variants or derivatives. In one embodiment of this type, the invention provides an isolated

polypeptide comprising at least biologically active fragment or portion of the sequence set forth in SEQ ID NO: 83, or a variant or derivative thereof.

Suitably, the biologically active fragment comprises at least one, and preferably both, of the consensus PPTase sequence motifs set forth in SEQ ID NO: 89 and 93, or variant thereof. Preferably, the biologically active fragment comprises the intervening sequence between the said consensus PPTase sequence motifs, which intervening sequence comprises the sequence set forth in SEQ ID NO: 91, or variant thereof.

Preferably, the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof.

Suitably, the variant has at least 60%, preferably at least 70%, more preferably at least 80%, more preferably at least 90% and still more preferably at least 95% sequence identity to the sequence set forth in SEQ ID NO: 83.

Preferably, the variant comprises at least one sequence selected from the group consisting of SEQ ID NO: 87, 89, 91 and 93, or variant thereof. In this regard, the variant preferably has at least 70%, preferably at least 80%, more preferably at least 90%, and still more preferably at least 95% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 87, 89, 91 or 93.

In yet another aspect, the present invention contemplates an isolated polypeptide encoding at least a portion of a methyltransferase (XabC) associated with albicidin biosynthesis or its variants or derivatives. In one embodiment of this type, the invention provides an isolated polypeptide comprising at least biologically active fragment or portion of the sequence set forth in SEQ ID NO: 95, or a variant or derivative thereof.

Suitably, the biologically active fragment comprises at least one, and preferably all, of the consensus methyltransferase sequence motifs set forth in SEQ ID NO: 99, 101 and 103, or variant thereof.

Preferably, the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 105, or variant thereof. In a preferred embodiment, the biologically active fragment comprises a contiguous

sequence of amino acids contained within the sequence set forth in SEQ ID NO: 107, or variant thereof.

Suitably, the variant has at least 60%, preferably at least 70%, more preferably at least 80%, more preferably at least 90% and still more preferably at least 95% sequence identity to the sequence set forth in SEQ ID NO: 95.

Preferably, the variant has at least 70%, preferably at least 80%, more preferably at least 90%, and still more preferably at least 95% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 99, 101 and 103.

In still yet another aspect, the invention contemplates an isolated polynucleotide encoding at least a portion of an albicidin PKS-NRPS (XabB) or its variants or derivatives, as broadly described above. Preferably, the polynucleotide comprises the sequence set forth in any one of SEQ ID NO: 1 and 3, or a biologically active fragment thereof, or a polynucleotide variant of these.

Suitably, the biologically active fragment is at least 18 nucleotides in length.

The polynucleotide preferably encodes at least one domain selected from the group consisting of the AL domain, the KS domain, the KR domain, the ACP domain, the A domain, the PCP domain and the C domain as broadly described above.

Suitably, the AL domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 5 and 7, or variants thereof. Preferably, the AL domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 5 and 7, or variants thereof.

The KS domain is preferably encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 9, 11, 13, 15, 17 and 19, or variants thereof. In one embodiment, the KS domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 9, 11 and 13, or variants thereof. In an alternate embodiment, the KS domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 15, 17 and 19, or variants thereof.

Preferably, the KR domain is encoded by a nucleotide sequence set forth in SEQ ID NO: 21, or variant thereof.

Suitably, the ACP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 23, 25 and 27, or variants thereof.

The A domain is preferably encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof. In a preferred embodiment, the A domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof.

Suitably, the PCP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 49 and 51, or variants thereof.

Preferably, the C domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79, or variants thereof. In one embodiment, the C domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 53, 55, 57, 59, 61, 63 and 65, or variants thereof. In an alternate embodiment, the C domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 67, 69, 71, 73, 75, 77 and 79, or variants thereof.

In one embodiment, the polynucleotide variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 1 or 3.

In another embodiment, the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 1 or 3 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

Preferably, the polynucleotide variant comprises a nucleotide sequence encoding at least one domain selected from the group consisting of the AL domain, the KS domain, the KR domain, the ACP domain, the A domain, the PCP domain and the C domain as broadly described above.

In one embodiment, the nucleotide sequence variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the sequences set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79.

In another embodiment, the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

In a further aspect, the invention contemplates an isolated polynucleotide encoding at least a portion of a PPTase (XabA) associated with albicidin biosynthesis or its variants or derivatives. an isolated polynucleotide encoding a polypeptide, fragment, variant or derivative as broadly described above. Preferably, the polynucleotide comprises the sequence set forth in any one of SEQ ID NO: 82 and 84, or a biologically active fragment thereof, or a polynucleotide variant of these.

Alternatively, the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 86, or variant thereof.

In one embodiment, the polynucleotide variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 82, 84 and 86.

In another embodiment, the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 82, 84 and 86 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

Preferably, the polynucleotide variant comprises a nucleotide sequence encoding at least one PPTase sequence motif selected from SEQ ID NO: 89 and 93, or variant thereof. Suitably, the polynucleotide variant comprises a nucleotide sequence encoding the

intervening sequence between the said consensus PPTase sequence motifs, said nucleotide sequence comprising the sequence set forth in SEQ ID NO: 91.

The polynucleotide variant suitably comprises a nucleotide sequence encoding a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof. In this instance, the contiguous sequence is preferably encoded by the sequence set forth in SEQ ID NO: 86, or nucleotide sequence variant thereof

Suitably, the PPTase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 88 and 92, or nucleotide sequence variant thereof.

Preferably, the said intervening sequence is encoded by the nucleotide sequence set forth in SEQ ID NO: 90, or nucleotide sequence variant thereof.

In one embodiment, the nucleotide sequence variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the sequences set forth in SEQ ID NO: 86, 88, 90 and 92.

In another embodiment, the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 86, 88, 90 and 92 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

In yet a further aspect, the invention contemplates an isolated polynucleotide encoding at least a portion of a methyltransferase (XabC) associated with albicidin biosynthesis or its variants or derivatives. Preferably, the polynucleotide comprises the sequence set forth in any one of SEQ ID NO: 94 and 96, or a biologically active fragment thereof, or a polynucleotide variant of these.

Alternatively the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 104, or variant thereof. In one embodiment, this polynucleotide preferably comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 106, or variant thereof

In one embodiment, the polynucleotide variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 94, 96, 104 and 106.

In another embodiment, the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 94, 96, 104 and 106 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

Preferably, the polynucleotide variant comprises a nucleotide sequence encoding a methyltransferase sequence motif selected from any one or more of SEQ ID NO: 99, 101 and 103, or variant thereof.

Suitably, the methyltransferase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 98, 100 and 102, or nucleotide sequence variant thereof.

In one embodiment, the nucleotide sequence variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the sequences set forth in SEQ ID NO: 98, 100 and 102.

In another embodiment, the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 98, 100 and 102 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

In still a further aspect, the invention features an expression vector comprising a polynucleotide as broadly described above wherein the polynucleotide is operably linked to a regulatory polynucleotide.

In another aspect, the invention provides a host cell containing a said expression vector.

Suitably, the host cell is a bacterium or other prokaryote.

In yet another aspect, the invention is directed to a multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression

vector for the production of a polypeptide, fragment, variant or derivative as broadly described above.

The invention also features a method of producing a recombinant polypeptide, fragment, variant or derivative as broadly described above, comprising:

– culturing a host cell containing an expression vector as broadly described above such that said recombinant polypeptide, fragment, variant or derivative is expressed from said polynucleotide; and

– isolating the said recombinant polypeptide, fragment, variant or derivative.

In another aspect, the invention provides a method of producing a biologically active fragment of a polypeptide as broadly described above, comprising:

– detecting an activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 2, wherein said activity is selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity and condensation activity; or

– detecting PPTase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 83; or

– detecting methyltransferase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 95;

wherein detection of said activity is indicative of said fragment being a biologically active fragment.

In a further aspect, the invention provides a method of producing a biologically active fragment as broadly described above, comprising:

– introducing a polynucleotide from which a fragment of a polypeptide as broadly described above can be produced into a cell; and

– detecting an activity selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity and condensation activity; or

– detecting PPTase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 83; or

– detecting methyltransferase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 95;

wherein detection of said activity is indicative of said fragment being a biologically active fragment.

In yet a further aspect, the invention provides a method of producing a variant of a polypeptide as broadly described above (parent polypeptide), or a biologically active fragment thereof, comprising:

– producing a modified polypeptide whose sequence is distinguished from the parent polypeptide or the biologically active fragment by substitution, deletion or addition of at least one amino acid; and

– detecting an activity associated with the modified polypeptide, wherein said activity is selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity, wherein detection of said activity is indicative of said modified polypeptide being a variant.

In a further aspect, the invention contemplates a method of producing a variant of a parent polypeptide as broadly described above, or biologically active fragment thereof, comprising:

– producing a polynucleotide from which a modified polypeptide as described above can be produced;

– introducing said polynucleotide into a cell; and

– detecting an activity associated with the modified polypeptide, wherein said activity is selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity, wherein detection of said activity is indicative of said modified polypeptide being a variant..

In yet another aspect, the invention extends to a method of screening for an agent that modulates the expression of a gene or variant thereof or the level and/or functional activity of an expression product of said gene or variant thereof, wherein said gene is

selected from *xabB*, *xabA*, or *xabC*, or a gene belonging to the same regulatory or biosynthetic pathway as *xabB*, *xabA*, or *xabC*, said method comprising:

– contacting a preparation comprising a polypeptide encoded by said gene, or biologically active fragment of said polypeptide, or variant or derivative of these, or a genetic sequence (*e.g.*, a transcriptional control element) that modulates the expression of said gene or variant thereof, with a test agent; and

– detecting a change in the level and/or functional activity of said polypeptide or biologically active fragment thereof, or variant or derivative, or of a product expressed from said genetic sequence.

The transcriptional control element preferably comprises the sequence set forth in SEQ ID NO: 81 or complement thereof.

The invention, in another aspect, also provides a method for enhancing the level and/or functional activity of an albicidin, said method comprising:

– introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of an albicidin PKS-NRPS or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of an albicidin PKS-NRPS or variant or derivative thereof can be translated;

– and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin.

Preferably, the method further comprises introducing into said host cell a vector from which a PPTase can be translated. Suitably, the PPTase is selected from EntD or XabA.

Preferably, the method further comprises introducing into said host cell a vector from which a methyltransferase, more preferably and *O*-methyltransferase, and even more preferably an *S*-adenosylmethionine *O*-methyltransferase can be translated.

According to another aspect of the invention, there is provided a method for enhancing the level and/or functional activity of an albicidin, said method comprising contacting a precursor of said albicidin or an intermediate involved in the biosynthesis of said albicidin with at least a portion of an albicidin PKS-NRPS, or variant or derivative

thereof, as broadly described above, for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin.

Preferably, the method further comprises contacting a precursor of said albicidin or an intermediate involved in the biosynthesis of said albicidin with a PPTase.

Preferably, the method further comprises contacting a precursor of said albicidin or an intermediate involved in the biosynthesis of said albicidin with a methyltransferase, more preferably and *O*-methyltransferase, and even more preferably an *S*-adenosylmethionine *O*-methyltransferase.

In another aspect, the invention provides a method of identifying a PPTase for enhancing the level and/or functional activity of an albicidin, said method comprising introducing into an albicidin-deficient strain of *X. albilineans* which lacks *xabA* a vector comprising a polynucleotide encoding a test PPTase, wherein said polynucleotide is operably linked to a regulatory polynucleotide, and detecting production of albicidin.

Suitably, the strain is LS156 described herein.

Preferably, the PPTase is EntD.

The invention, in another aspect, also provides a method for enhancing the level and/or functional activity of an albicidin, said method comprising:

- introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of a PPTase associated with albicidin biosynthesis or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of a PPTase associated with albicidin biosynthesis or variant or derivative thereof can be translated;
- and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin

In yet another aspect, the invention provides a method for enhancing the level and/or functional activity of an albicidin, said method comprising:

- introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of a methyltransferase associated

with albicidin biosynthesis or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of a methyltransferase associated with albicidin biosynthesis or variant or derivative thereof can be translated;

– and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin

In another aspect, the invention resides in an antigen-binding molecule that is immuno-interactive with a polypeptide, fragment, variant or derivative as broadly described above.

In yet another aspect, the invention provides a method to prepare a polynucleotide encoding a modified PKS, comprising using an albicidin PKS-NRPS encoding nucleotide sequence as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement.

In still yet another aspect, the invention contemplates a method for producing polyketides, comprising expressing the modified albicidin PKS encoding nucleotide sequence as broadly described in a suitable host cell to thereby produce a polyketide different from that produced by the albicidin PKS-NRPS.

Another aspect of the invention contemplates the insertion of portions of the albicidin PKS-NRPS coding sequence into other PKS coding sequences to modify the products thereof.

In a further aspect, the invention encompasses use of the polypeptide, fragment, variant or derivative as broadly described above, or the polynucleotide or vector as broadly described above, or the modulatory agent as broadly described above for producing secondary metabolites, preferably albicidins.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation showing a physical and functional map of part of the albicidin biosynthetic gene cluster. (A). Partial physical map of the Tn5 insertion locus in LS157 genomic DNA. Restriction enzymes used: C, *Cla*I; E, *Eco*RI; S, *Spe*I; N, *Not*I; and B, *Bam*HI. (B). Probes used to recover clone pXABB: Probe 1, 1.4-kb *Eco*RI-*Not*I fragment digested from pBC157; and probe 2, 0.9-kb PCR product amplified from Xa13 genomic DNA using primers complementary to sequences flanking the Tn5 insertion in LS157. (C). Clones and subclones used for sequencing, and described in Table 1. (D). The transcription directions of three putative ORFs in 16.5-kb *Eco*RI fragment are indicated by arrows. (E). Organisation of *X. albilineans* XabB constructed by comparison with known protein sequences. The unshaded box indicates PKS region, and the shade box indicates NRPS region. Relative positions of potential catalytic domains or active sites are indicated by: AL, acyl-CoA ligase; ACP, acyl carrier protein; KS, β-ketoacyl synthase; KR, β-ketoacyl reductase; PCP, peptidyl carrier protein; C, condensation; A, adenylation. Horizontal bars indicate proposed biosynthetic modules.

Figure 2 is a diagrammatic representation presenting the sequence of the region upstream from *xabB*. The nucleotide sequence is numbered according to the 16511-bp sequence in GenBank accession no. AF239749. The putative –35 and –10 promoter sequences of *xabB* and the divergent gene *xatA* are underlined, as are ribosome-binding sequences. The transcriptional directions of *xabB* and *xatA* are indicated by arrows. Translational start codons are indicated by boldface type. Primers P1F1 and P1R are shaded.

Figure 3 is a diagrammatic representation showing the alignment of *X. albilineans* XabB enzymatic domains with those of PKSs and FASs from other organisms. Identical amino acids are indicated by boldface type. Stars and overlines identify conserved amino acids at catalytic sites. Xal-XabB, *X. albilineans* XabB for biosynthesis of albicidin (this study); Hin-LCFA, *Haemophilus influenza* long-chain fatty acid-CoA ligase (P46450); Bsu-PksJ, *B. subtilis* polyketide synthase J (P40806); Bsu-MycA, *B. subtilis* MycA for biosynthesis of mycosubtilin (AF184956); Pcr-ComL2, *Petroselinum crispum* 4-coumarate-CoA ligase 2 (P14913); Sma-FkbB, *S.* sp. MA6548 FkbB for biosynthesis of FK506 (AF082099); Ame-RifA, *Amycolatopsis mediterranei* RifA for biosynthesis of

rifamycin B (AF040570); Shy-RapA, *S. hygroscopicus* RapA for biosynthesis of rapamycin (X86780); Mxa-Ta1, *M. xanthus* Ta1 for biosynthesis of TA (AJ006977); Ser-EryA1 and EryA3, *S. erythraea* EryA modules for biosynthesis of erythromycin (M63676, M63677); Che-PKS1, *Cochliobolus heterostrophus* PKS1 for biosynthesis of T-toxin (U68040); Bsu-PksM, *B. subtilis* PKS for a polyketide synthase (O31781); Mtu-PpsA, *M. tuberculosis* PKS for a polyketide synthase (G3261605); Mtu-MAS, *M. tuberculosis* MAS for biosynthesis of mycocerosic acid (M95808); Chick-FAS, chichen fatty acid synthase (M22987); Rat-FAS, rat fatty acid synthase (X14175).

Figure 4 is a graphical representation showing albicidin production by wild-type *X. albilineans* LS155 (▲), complemented Tox$^-$ mutant strain LS157 pLXABB1 (O), complemented Tox$^-$ mutant strain LS157 pLXABB2 (●), LS157 (■), and LS157 pLAFR3 (+). Albicidin concentrations in culture supernatants were quantified based on inhibition zone width in a microbial bioassay (means +/- standard errors from 5 replicates).

Figure 5 is a graphical representation showing the relationship between growth (■), albicidin production (O), and GUS activity (▲) in *X. albilineans* LS155 pRG960p1 (A) and in LS155 pRG960p2 (B). Relative activity (means +/- standard errors from 2 replicates): 100% growth, OD_{550} = 1.43; 100% albicidin production = 268.5 units/ml; 100% GUS activity = 119 units/mg of protein (one unit equals 1 pmol of methylumbelliferone formed per min.). Locations and sizes of inserts on pRG960p1 and pRG960p2 are indicated in Figure 2 and Table 1. GUS, β-glucuronidase.

Figure 6 is a schematic representation showing the organisation of five known PKS-NRPS enzymes. *X. albilineans* XabB, encoded by *xabB* for albicidin biosynthesis (this study); *B. subtilis* MycA for mycosubtilin biosynthesis (Duitman *et al.*, 1999); *Yersinia pestis* HMWP1 for yersiniabactin biosynthesis (Gehring *et al.*, 1998); *M. xanthus* partial gene product Ta1 for TA biosynthesis (Paitan *et al.*, 1999); *B. subtilis* PksorfX6 for unknown function (Albertini *et al.*, 1995). Unshaded boxes indicate PKS regions, grey boxes indicate NRPS regions, and dark boxes indicate amino transferase (AMT) or methyltransferase (MT). Vertical bars follow the carrier domains at the end of each biosynthetic "module".

Figure 7 is a diagrammatic representation showing a dendrogram (GCG) analysis of adenylation domains of XabB and its homologous peptide synthetases. Peptide

synthetases, including various modules of the same multienzyme complex, are as follows: GrsA and GrsB, gramicidin synthetase A and B, respectively, from *B. subtilis* (X15577, X61658); BacA, BacB, and BacC, bacitracin synthetase A, B, and C, respectively, from *B. licheniformis* (AF007865); SnbC and SnbDE, pristinamycin I synthetase C and DE, respectively, from *S. pristinaespiralis* (X98690, Y11547); FkbP, FK506 synthetase FkbP from *S.* sp. MA6548 (AF082100); TycA, TycB, and TycC, tyrocidine synthetase A, B, and C, respectively, from *B. brevis* (AF004835); SyrE, syringomycin synthetase E1 from *Pseudomonas syringae* pv. *syringae* (AF047828); EntF, enterobactin synthetase F from *E. coli* (P11454); DhbF, 2,3-dihydroxybenzoate synthetase F from *B. subtilis* (P45745); FenD, fengycin synthetase FenD1 from *B. subtilis* (AJ011849); SrfAA, SrfAB, and SrfAC, surfactin A synthetase A, B, and C, respectively, from *B. subtilus* (X70356); XabB, albicidin synthase B from *X. albilineans* (this study). The A4 to A5 regions (about 100 aa) of adenylation domains of peptide synthetases, which is involved in amino acid recognition and binding, were aligned using the PILEUP program with default parameters.

Figure 8 is a diagrammatic representation showing a restriction map of clones including the *xabA* gene from *X. albilineans*. Sequencing by primer walking commenced at the T3 and T7 primers. The location and direction of transcription of the *xabA* ORF is shown by an arrow. Restriction enzymes are: E, *Eco*RI; P, *Pst*I; C, *Cla*I; and H, *Hin*dIII

Figure 9 is a diagrammatic representation presenting the sequence of the *xabA* gene. The nucleotide sequence is numbered according to the 3-kb sequence in GenBank accession no. AF191324. The closest matches to RBS region and promoter consensus sequences are underlined, as are the region of dyad symmetry and putative factor-independent termination sites. Translation start and stop codons are indicated by boldface type. The (V/I)G(V/I)D and (F/W)(S/C/T)xKE(A/S)xxK motifs conserved in PPTase enzymes are boxed. The insertion site of Tn5 is marked (▼).

Figure 10 is a graphical representation showing albicidin production by wild-type *X. albilineans* strain Xa13 (O), Xa13 pLXABA (•), and complemented Tox$^-$ mutant strain LS156 pLXABA (▲). Albicidin concentrations in culture supernatants were quantified based on inhibition zone width in a microbial bioassay (means +/- standard errors from 2 replicates).

Figure 11 is a schematic representation showing a dendrogram (GCG) analysis of PPTases involved in antibiotic and fatty acid biosynthesis in bacteria. Sau, *Salmonella austin*; Sty, *Salmonella typhymurium*; Bbr, *Bacillus brevis*; Xal, *Xanthomonas albilineans*; Eco, *Escherichia coli*; Sfl, *Shigella flexneri*; Bpu, *Bacillus pumilus*; Bsu, *Bacillus subtilis*; Mtu, *Mycobacterium tuberculosis*; Hin, *Haemophilus influenzae*. The sources of amino acid sequence of PPTases correspond to those in Table 2, and the sequences were aligned using the PILEUP program with default parameters.

Figure 12 is a schematic representation showing the organisation of part of the albicidin biosynthetic gene cluster. The location and direction of three ORFs are indicated by thick arrows. Vertical lines indicate the position of restriction enzyme sites: E, *Eco*RI; B, *Bam*HI; S, *Spe*I; N, *Nco*I. The vertical lines with triangles (▲) show the position of insertional mutagenesis sites or Tn5 insertion site, and the resultant mutants are bracketed. The arrows above the physical map indicate the locations of primers used to amplify sequence downstream of the *Eco*RI restriction site by IPCR. The cloned regions for complementation tests are shown below the map.

Figure 13 is a diagrammatic representation presenting the nucleotide and deduced amino acid sequences of the *xabC* region. The nucleotide sequence is numbered according to the 1515-bp sequence in GenBank accession no. AF239750. The potential RBS and selected restriction sites are underlined. The putative factor-independent termination signals are underlined and indicated by bold letters. Translation start and stop codons are indicated by bold letters. The conserved motifs in Mtases are boxed. Primers used for PCR (A3F and A3R) and IPCR (IR) are shaded.

Figure 14 is a diagrammatic representation showing the conserved sequence motifs in Mtases involved in antibiotic biosynthesis in bacteria. Identical or similar amino acids (A = G; D = E; I = L = V) are shown in bold. Numbers indicate amino acid residues from the N terminus of the protein. Xal-XabC, putative albicidin biosynthesis Mtase from *X. albilineans* (this study); Sgl-TcmO and Sgl-TcmN, multifunctional cyclase-dehydrase-3-*O*-Mtase and tetracenomycin polyketide synthesis 8-*O*-Mtase of *Streptomyces glaucescens*, respectively (accession number M80674); Smy-MdmC, midecamycin-*O*-Mtase of *S. mycarofaciens* (M93958); Mxa-SafC, saframycin *O*-Mtase of *Myxococcus xanthus* (U24657); Ser-EryG, erythromycin biosynthesis *O*-Mtase of *Saccharopolyspora*

erythraea (S18533); Spe-DauK, carminomycin 4-*O*-Mtase from *S. peucetius* (L13453); Sal-DmpM, *O*-demethylpuromycin-*O*-Mtase from *S. alboniger* (M74560); Shy-RapM, rapamycin *O*-Mtase of *S. hygroscopicus* (X86780); Sav-AveD, avermectin B 5-*O*-Mtase from *S. avermitilis* (G5921167).

Figure 15 is a graphical representation showing albicidin production by wild-type *X. albilineans* LS155 (○), Tox⁻ *xabC* insertion mutant LS-JP2 (□), complemented strain LS-JP2 pLXABC containing Lac promoter – full length *xabC* gene (O), and complemented strain LS-JP2 pLXABB1 containing full length *xabB* plus functional N-terminal region of *xabC* (∏). Albicidin concentrations in culture supernatants were quantified based on inhibition zone width in a microbial bioassay (means +/- standard errors from 2 or 3 replicates).

BRIEF DESCRIPTION OF THE SEQUENCES: SUMMARY TABLE

TABLE A

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 1	Full-length *xabB* (Accession No. AF239749)	16551 bases
SEQ ID NO: 2	Full-length polypeptide sequence encoded by SEQ ID NO: 1	4801 residues
SEQ ID NO: 3	Full-length coding sequence of *xabB*	14406 bases
SEQ ID NO: 4	Polypeptide sequence encoded by SEQ ID NO: 3	4801 residues
SEQ ID NO: 5	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl-CoA ligase subdomain I	45 bases
SEQ ID NO: 6	Acyl-CoA ligase subdomain I encoded by SEQ ID NO: 5	15 residues
SEQ ID NO: 7	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl-CoA ligase subdomain II	24 bases
SEQ ID NO: 8	Acyl-CoA ligase subdomain I encoded by SEQ ID NO: 7	8 residues
SEQ ID NO: 9	Sub-sequence of SEQ ID NO: 1 and 3 encoding *β*-ketoacyl synthase 1 subdomain I	51 bases
SEQ ID NO: 10	*β*-Ketoacyl synthase 1 subdomain I encoded by SEQ ID NO: 9	17 residues
SEQ ID NO: 11	Sub-sequence of SEQ ID NO: 1 and 3 encoding *β*-ketoacyl synthase 1 subdomain II	30 bases
SEQ ID NO: 12	*β*-Ketoacyl synthase 1 subdomain II encoded by SEQ ID NO: 11	10 residues
SEQ ID NO: 13	Sub-sequence of SEQ ID NO: 1 and 3 encoding *β*-ketoacyl synthase 1 subdomain III	30 bases
SEQ ID NO: 14	*β*-Ketoacyl synthase 1 subdomain III encoded by SEQ ID NO: 13	10 residues
SEQ ID NO: 15	Sub-sequence of SEQ ID NO: 1 and 3 encoding *β*-ketoacyl synthase 2 subdomain I	51 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 16	β-Ketoacyl synthase 2 subdomain I encoded by SEQ ID NO: 15	17 residues
SEQ ID NO: 17	Sub-sequence of SEQ ID NO: 1 and 3 encoding β-ketoacyl synthase 2 subdomain II	30 bases
SEQ ID NO: 18	β-Ketoacyl synthase 2 subdomain II encoded by SEQ ID NO: 17	10 residues
SEQ ID NO: 19	Sub-sequence of SEQ ID NO: 1 and 3 encoding β-ketoacyl synthase 2 subdomain III	30 bases
SEQ ID NO: 20	β-Ketoacyl synthase 2 subdomain III encoded by SEQ ID NO: 19	10 residues
SEQ ID NO: 21	Sub-sequence of SEQ ID NO: 1 and 3 encoding β-ketoacyl reductase domain	93 bases
SEQ ID NO: 22	β-Ketoacyl reductase domain encoded by SEQ ID NO: 21	31 residues
SEQ ID NO: 23	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl carrier protein 1 domain	36 bases
SEQ ID NO: 24	Acyl carrier protein 1 domain encoded by SEQ ID NO: 23	12 residues
SEQ ID NO: 25	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl carrier protein 2 domain	36 bases
SEQ ID NO: 26	Acyl carrier protein 2 domain encoded by SEQ ID NO: 25	12 residues
SEQ ID NO: 27	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl carrier protein 3 domain	36 bases
SEQ ID NO: 28	Acyl carrier protein 3 domain encoded by SEQ ID NO: 27	12 residues
SEQ ID NO: 29	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain I	18 bases
SEQ ID NO: 30	Adenylation domain subdomain I encoded by SEQ ID NO: 29	6 residues
SEQ ID NO: 31	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain II	33 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 32	Adenylation domain subdomain II encoded by SEQ ID NO: 31	11 residues
SEQ ID NO: 33	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain III	48 bases
SEQ ID NO: 34	Adenylation domain subdomain III encoded by SEQ ID NO: 33	16 residues
SEQ ID NO: 35	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain IV	12 bases
SEQ ID NO: 36	Adenylation domain subdomain IV encoded by SEQ ID NO: 35	4 residues
SEQ ID NO: 37	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain V	21 bases
SEQ ID NO: 38	Adenylation domain subdomain V encoded by SEQ ID NO: 37	7 residues
SEQ ID NO: 39	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain VI	45 bases
SEQ ID NO: 40	Adenylation domain subdomain VI encoded by SEQ ID NO: 39	15 residues
SEQ ID NO: 41	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain VII	18 bases
SEQ ID NO: 42	Adenylation domain subdomain VII encoded by SEQ ID NO: 41	6 residues
SEQ ID NO: 43	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain VIII	60 bases
SEQ ID NO: 44	Adenylation domain subdomain VIII encoded by SEQ ID NO: 43	20 residues
SEQ ID NO: 45	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain IX	21 bases
SEQ ID NO: 46	Adenylation domain subdomain IX encoded by SEQ ID NO: 45	7 residues
SEQ ID NO: 47	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain X	18 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 48	Adenylation domain subdomain X encoded by SEQ ID NO: 47	6 residues
SEQ ID NO: 49	Sub-sequence of SEQ ID NO: 1 and 3 encoding peptidyl carrier protein 1 domain	33 bases
SEQ ID NO: 50	Peptidyl carrier protein 1 domain encoded by SEQ ID NO: 49	11 residues
SEQ ID NO: 51	Sub-sequence of SEQ ID NO: 1 and 3 encoding peptidyl carrier protein 2 domain	33 bases
SEQ ID NO: 52	Peptidyl carrier protein 2 domain encoded by SEQ ID NO: 51	11 residues
SEQ ID NO: 53	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain I	30 bases
SEQ ID NO: 54	Condensation domain 1 subdomain I encoded by SEQ ID NO: 53	10 residues
SEQ ID NO: 55	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain II	27 bases
SEQ ID NO: 56	Condensation domain 1 subdomain II encoded by SEQ ID NO: 55	9 residues
SEQ ID NO: 57	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain III	30 bases
SEQ ID NO: 58	Condensation domain 1 subdomain III encoded by SEQ ID NO: 57	10 residues
SEQ ID NO: 59	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain IV	21 bases
SEQ ID NO: 60	Condensation domain 1 subdomain IV encoded by SEQ ID NO: 59	7 residues
SEQ ID NO: 61	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain V	36 bases
SEQ ID NO: 62	Condensation domain 1 subdomain V encoded by SEQ ID NO: 61	12 residues
SEQ ID NO: 63	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain VI	21 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 64	Condensation domain 1 subdomain VI encoded by SEQ ID NO: 63	7 residues
SEQ ID NO: 65	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain VII	24 bases
SEQ ID NO: 66	Condensation domain 1 subdomain VII encoded by SEQ ID NO: 65	8 residues
SEQ ID NO: 67	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain I	30 bases
SEQ ID NO: 68	Condensation domain 2 subdomain I encoded by SEQ ID NO: 67	10 residues
SEQ ID NO: 69	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain II	27 bases
SEQ ID NO: 70	Condensation domain 2 subdomain II encoded by SEQ ID NO: 69	9 residues
SEQ ID NO: 71	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain III	30 bases
SEQ ID NO: 72	Condensation domain 2 subdomain III encoded by SEQ ID NO: 71	10 residues
SEQ ID NO: 73	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain IV	21 bases
SEQ ID NO: 74	Condensation domain 2 subdomain IV encoded by SEQ ID NO: 73	7 residues
SEQ ID NO: 75	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain V	33 bases
SEQ ID NO: 76	Condensation domain 2 subdomain V encoded by SEQ ID NO: 75	11 residues
SEQ ID NO: 77	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain VI	21 bases
SEQ ID NO: 78	Condensation domain 2 subdomain VI encoded by SEQ ID NO: 77	7 residues
SEQ ID NO: 79	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain VII	24 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 80	Condensation domain 2 subdomain VII encoded by SEQ ID NO: 79	8 residues
SEQ ID NO: 81	Polynucleotide comprising *xabB* promoter	242 bases
SEQ ID NO: 82	Full-length *xabA* (Accession No. AF191324)	1200 bases
SEQ ID NO: 83	Full-length polypeptide sequence encoded by SEQ ID NO: 82	278 residues
SEQ ID NO: 84	Full-length coding sequence of *xabA*	837 bases
SEQ ID NO: 85	Polypeptide sequence encoded by SEQ ID NO: 84	278 residues
SEQ ID NO: 86	Sub-sequence of SEQ ID NO: 82 encoding PPTase domain	168 bases
SEQ ID NO: 87	PPTase domain encoded by SEQ ID NO: 86	56 residues
SEQ ID NO: 88	Sub-sequence of SEQ ID NO: 82 encoding a motif (motif I) conserved in PPTases	27 bases
SEQ ID NO: 89	PPTase motif I amino acid sequence encoded by SEQ ID NO: 88	9 residues
SEQ ID NO: 90	Sub-sequence of SEQ ID NO: 82 encoding intervening amino acid sequence linking motifs I and II	117 bases
SEQ ID NO: 91	Intervening amino acid sequence encoded by SEQ ID NO: 90	39 residues
SEQ ID NO: 92	Sub-sequence of SEQ ID NO: 82 encoding a motif (motif II) conserved in PPTases	36 bases
SEQ ID NO: 93	PPTase motif II amino acid sequence encoded by SEQ ID NO: 92	12 residues
SEQ ID NO: 94	Full-length *xabC* (Accession No. AF239750)	1515 bases
SEQ ID NO: 95	Full-length polypeptide sequence encoded by SEQ ID NO: 94	343 residues
SEQ ID NO: 96	Full-length coding sequence of *xabC*	1029 bases
SEQ ID NO: 97	Polypeptide sequence encoded by SEQ ID NO: 96	343 residues

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 98	Sub-sequence of SEQ ID NO: 94 encoding a motif (motif I) conserved in methyltransferases	21 bases
SEQ ID NO: 99	Methyltransferase motif I amino acid sequence encoded by SEQ ID NO: 98	7 residues
SEQ ID NO: 100	Sub-sequence of SEQ ID NO: 94 encoding a motif (motif II) conserved in methyltransferases	24 bases
SEQ ID NO: 101	Methyltransferase motif II amino acid sequence encoded by SEQ ID NO: 100	8 residues
SEQ ID NO: 102	Sub-sequence of SEQ ID NO: 94 encoding a motif (motif III) conserved in methyltransferases	27 bases
SEQ ID NO: 103	Methyltransferase motif III amino acid sequence encoded by SEQ ID NO: 102	9 residues
SEQ ID NO: 104	Polynucleotide encoding said motifs I, II and III	303 bases
SEQ ID NO: 105	Polypeptide encoded by SEQ ID NO: 104	101 residues
SEQ ID NO: 106	Biologically active fragment of SEQ ID NO: 94	831 bases
SEQ ID NO: 107	Biologically active fragment of SEQ ID NO: 95	277 residues

DETAILED DESCRIPTION OF THE INVENTION

1. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

The articles *"a"* and *"an"* are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The term *"about"* is used herein to refer to sequences that vary by as much as 30%, preferably by as much as 20% and more preferably by as much as 10% to the length of a reference sequence.

By *"agent"* is meant a naturally occurring or synthetically produced molecule which interacts either directly or indirectly with a target member, the level and/or functional activity of which are to be modulated.

"Amplification product" refers to a nucleic acid product generated by nucleic acid amplification techniques.

By *"antigen-binding molecule"* is meant a molecule that has binding affinity for a target antigen. It will be understood that this term extends to immunoglobulins, immunoglobulin fragments and non-immunoglobulin derived protein frameworks that exhibit antigen-binding activity.

As used herein, the term *"binds specifically"* and the like refers to antigen-binding molecules that bind the polypeptide or polypeptide fragments of the invention but do not significantly bind to homologous prior art polypeptides.

By *"biologically active fragment"* is meant a fragment of a full-length parent polypeptide which fragment retains the activity of the parent polypeptide. A biologically active fragment will therefore comprise an activity selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity. As used herein, the term *"biologically active fragment"* includes deletion mutants and small peptides, for example of at least 10, preferably at least 20 and more preferably at least 30 contiguous amino acids, which comprise the above activities. Peptides of this type may be obtained through the application of standard recombinant nucleic acid techniques or synthesised using conventional liquid or solid phase synthesis techniques. For example, reference may be made to solution synthesis or solid phase synthesis as described, for example, in Chapter 9 entitled *"Peptide Synthesis"* by Atherton and Shephard which is included in a publication entitled *"Synthetic Vaccines"* edited by Nicholson and published by Blackwell Scientific Publications. Alternatively, peptides can be produced by digestion of a polypeptide of the invention with proteinases such as endoLys-C, endoArg-C, endoGlu-C and staphylococcus V8-protease. The digested fragments can be purified by, for example, high performance liquid chromatographic (HPLC) techniques.

Throughout this specification, unless the context requires otherwise, the words *"comprise"*, *"comprises"* and *"comprising"* will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

By "*corresponds to*" or "*corresponding to*" is meant a polynucleotide (a) having a nucleotide sequence that is substantially identical or complementary to all or a portion of a reference polynucleotide sequence or (b) encoding an amino acid sequence identical to an amino acid sequence in a peptide or protein. This phrase also includes within its scope a peptide or polypeptide having an amino acid sequence that is substantially identical to a sequence of amino acids in a reference peptide or protein.

By *"derivative"* is meant a polypeptide that has been derived from the basic sequence by modification, for example by conjugation or complexing with other chemical moieties or by post-translational modification techniques as would be understood in the art.

The term "*derivative*" also includes within its scope alterations that have been made to a parent sequence including additions, or deletions that provide for functionally equivalent molecules. Accordingly, the term derivative encompasses molecules that will have an activity selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity.

"*Homology*" refers to the percentage number of amino acids that are identical or constitute conservative substitutions as defined in Table B *infra*. Homology may be determined using sequence comparison programs such as GAP (Deveraux *et al.* 1984, *Nucleic Acids Research* **12**, 387-395). In this way, sequences of a similar or substantially different length to those cited herein might be compared by insertion of gaps into the alignment, such gaps being determined, for example, by the comparison algorithm used by GAP.

"*Hybridisation*" is used herein to denote the pairing of complementary nucleotide sequences to produce a DNA-DNA hybrid or a DNA-RNA hybrid. Complementary base sequences are those sequences that are related by the base-pairing rules. In DNA, A pairs with T and C pairs with G. In RNA U pairs with A and C pairs with G. In this regard, the terms "match" and "mismatch" as used herein refer to the hybridisation potential of paired nucleotides in complementary nucleic acid strands. Matched nucleotides hybridise efficiently, such as the classical A-T and G-C base pair mentioned above. Mismatches are other combinations of nucleotides that do not hybridise efficiently.

Reference herein to "*immuno-interactive*" includes reference to any interaction, reaction, or other form of association between molecules and in particular where one of the molecules is, or mimics, a component of the immune system.

By "*immuno-interactive fragment*" is meant a fragment of a parent or reference polypeptide as described herein, which fragment elicits an immune response, including the production of elements that specifically bind to said polypeptide, or variant or derivative thereof. As used herein, the term "*immuno-interactive fragment*" includes deletion mutants and small peptides, for example of at least six, preferably at least 8 and more preferably at

least 20 contiguous amino acids, which comprise antigenic determinants or epitopes. Several such fragments may be joined together.

By *"isolated"* is meant material that is substantially or essentially free from components that normally accompany it in its native state. For example, an "isolated polynucleotide", as used herein, refers to a polynucleotide, which has been purified from the sequences which flank it in a naturally occurring state, *e.g.*, a DNA fragment which has been removed from the sequences which are normally adjacent to the fragment.

By *"modulating"* is meant increasing or decreasing, either directly or indirectly, the level and/or functional activity of a target molecule. For example, an agent may indirectly modulate the said level/activity by interacting with a molecule other than the target molecule. In this regard, indirect modulation of a gene encoding a target polypeptide includes within its scope modulation of the expression of a first nucleic acid molecule, wherein an expression product of the first nucleic acid molecule modulates the expression of a nucleic acid molecule encoding the target polypeptide.

By *"obtained from"* is meant that a sample such as, for example, a nucleic acid extract or polypeptide extract is isolated from, or derived from, a particular source. For example, the extract may be isolated directly from any organism that produces secondary metabolites, preferably from an albicidin-producing microorganism, more preferably from microorganisms of the genus *Xanthomonas*.

The term "*oligonucleotide*" as used herein refers to a polymer composed of a multiplicity of nucleotide units (deoxyribonucleotides or ribonucleotides, or related structural variants or synthetic analogues thereof) linked via phosphodiester bonds (or related structural variants or synthetic analogues thereof). Thus, while the term "oligonucleotide" typically refers to a nucleotide polymer in which the nucleotides and linkages between them are naturally occurring, it will be understood that the term also includes within its scope various analogues including, but not restricted to, peptide nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. The exact size of the molecule may vary depending on the particular application. An oligonucleotide is typically rather short in length, generally from about 10 to 30 nucleotides, but the term can refer to molecules of any length, although the term "polynucleotide" or "nucleic acid" is typically used for large oligonucleotides.

By *"operably linked"* is meant that transcriptional and translational regulatory nucleic acids are positioned relative to a polypeptide-encoding polynucleotide in such a manner that the polynucleotide is transcribed and the polypeptide is translated.

The term *"polynucleotide"* or *"nucleic acid"* as used herein designates mRNA, RNA, cRNA, cDNA or DNA. The term typically refers to oligonucleotides greater than 30 nucleotides in length.

The terms *"polynucleotide variant"* and *"variant"* refer to polynucleotides displaying substantial sequence identity with a reference polynucleotide sequence or polynucleotides that hybridise with a reference sequence under stringent conditions that are defined hereinafter. These terms also encompass polynucleotides in which one or more nucleotides have been added or deleted, or replaced with different nucleotides. In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference polynucleotide whereby the altered polynucleotide retains the biological function or activity of the reference polynucleotide. The terms *"polynucleotide variant"* and *"variant"* also include naturally occurring allelic variants.

"Polypeptide", *"peptide"* and *"protein"* are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers.

The term *"polypeptide variant"* refers to polypeptides in which one or more amino acids have been replaced by different amino acids. It is well understood in the art that some amino acids may be changed to others with broadly similar properties without changing the nature of the activity of the polypeptide (conservative substitutions) as described hereinafter. These terms also encompass polypeptides in which one or more amino acids have been added or deleted, or replaced with different amino acids. Accordingly, polypeptide variants as used herein encompass polypeptides that have an activity selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl

carrier protein activity, condensation activity, PPTase activity and methyltransferase activity.

By *"primer"* is meant an oligonucleotide which, when paired with a strand of DNA, is capable of initiating the synthesis of a primer extension product in the presence of a suitable polymerising agent. The primer is preferably single-stranded for maximum efficiency in amplification but may alternatively be double-stranded. A primer must be sufficiently long to prime the synthesis of extension products in the presence of the polymerisation agent. The length of the primer depends on many factors, including application, temperature to be employed, template reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 or more nucleotides, although it may contain fewer nucleotides. Primers can be large polynucleotides, such as from about 200 nucleotides to several kilobases or more. Primers may be selected to be "substantially complementary" to the sequence on the template to which it is designed to hybridise and serve as a site for the initiation of synthesis. By "substantially complementary", it is meant that the primer is sufficiently complementary to hybridise with a target nucleotide sequence. Preferably, the primer contains no mismatches with the template to which it is designed to hybridise but this is not essential. For example, non-complementary nucleotides may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the template. Alternatively, non-complementary nucleotides or a stretch of non-complementary nucleotides can be interspersed into a primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridise therewith and thereby form a template for synthesis of the extension product of the primer.

"Probe" refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another molecule. Unless otherwise indicated, the term "probe" typically refers to a polynucleotide probe that binds to another nucleic acid, often called the "target nucleic acid", through complementary base pairing. Probes may bind target nucleic acids lacking complete sequence complementarity with the probe, depending on the stringency of the hybridisation conditions. Probes can be labelled directly or indirectly.

The term *"recombinant polynucleotide"* as used herein refers to a polynucleotide formed *in vitro* by the manipulation of nucleic acid into a form not normally found in nature. For example, the recombinant polynucleotide may be in the form of an expression vector. Generally, such expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleotide sequence.

By *"recombinant polypeptide"* is meant a polypeptide made using recombinant techniques, *i.e.*, through the expression of a recombinant polynucleotide.

By *"reporter molecule"* as used in the present specification is meant a molecule that, by its chemical nature, provides an analytically identifiable signal that allows the detection of a complex comprising an antigen-binding molecule and its target antigen. The term "reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity" and "substantial identity". A "*reference sequence*" is at least 12 but frequently 15 to 18 and often at least 25 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (*i.e.*, only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "*comparison window*" refers to a conceptual segment of at least 6 contiguous positions, usually about 50 to about 100, more usually about 100 to about 150 in which a sequence is compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerised implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release

7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (*i.e.*, resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons Inc, 1994-1998, Chapter 15.

The term *"sequence identity"* as used herein refers to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a *"percentage of sequence identity"* is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) or the identical amino acid residue (*e.g.*, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, *"sequence identity"* will be understood to mean the "match percentage" calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA) using standard defaults as used in the reference manual accompanying the software.

"Stringency" as used herein, refers to the temperature and ionic strength conditions, and presence or absence of certain organic solvents, during hybridisation and washing procedures. The higher the stringency, the higher will be the degree of complementarity between immobilised target nucleotide sequences and the labelled probe polynucleotide sequences that remain hybridised to the target after washing.

"Stringent conditions" refers to temperature and ionic conditions under which only nucleotide sequences having a high frequency of complementary bases will hybridise. The stringency required is nucleotide sequence dependent and depends upon the various components present during hybridisation and subsequent washes, and the time allowed for these processes. Generally, in order to maximise the hybridisation rate, non-stringent

hybridisation conditions are selected; about 20 to 25 °C lower than the thermal melting point (T_m). The T_m is the temperature at which 50% of specific target sequence hybridises to a perfectly complementary probe in solution at a defined ionic strength and pH. Generally, in order to require at least about 85% nucleotide complementarity of hybridised sequences, highly stringent washing conditions are selected to be about 5 to 15 °C lower than the T_m. In order to require at least about 70% nucleotide complementarity of hybridised sequences, moderately stringent washing conditions are selected to be about 15 to 30 °C lower than the T_m. Highly permissive (low stringency) washing conditions may be as low as 50 °C below the T_m, allowing a high level of mis-matching between hybridised sequences. Those skilled in the art will recognise that other physical and chemical parameters in the hybridisation and wash stages can also be altered to affect the outcome of a detectable hybridisation signal from a specific level of homology between target and probe sequences. Other examples of stringency conditions are described in section 3.3.

By *"vector"* is meant a nucleic acid molecule, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, or plant virus, into which a nucleic acid sequence may be inserted or cloned. A vector preferably contains one or more unique restriction sites and may be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector may be an autonomously replicating vector, *i.e.*, a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a linear or closed circular plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into a cell, is integrated into the genome of the recipient cell and replicated together with the chromosome(s) into which it has been integrated. A vector system may comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the cell into which the vector is to be introduced. The vector may also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are well known to those of skill in the art.

As used herein, underscoring or italicising the name of a gene shall indicate the gene, in contrast to its protein product, which is indicated by the name of the gene in the absence of any underscoring or italicising. For example, "*xabB*" shall mean the *xabB* gene, whereas "XabB" shall indicate the protein product of the "*xabB*" gene.

2. *Isolated polypeptides, biologically active fragments, polypeptide variants and derivatives*

2.1 Polypeptides of the invention

2.1.1 Albicidin synthetase

The present inventor has also isolated a gene (*xabB*) encoding a large modular polyketide synthase (PKS) linked to a non-ribosomal peptide synthetase (NRPS) (predicted *Mr* 525,695). At 4801 amino acids in length, the product of *xabB* (XabB) is the largest reported PKS-NRPS. Comparison of XabB with available protein sequence databases reveals an N-terminal region (from Met-1 to Asp-3235) similar to many microbial modular PKSs, and a C-terminal region (from Pro-3236 to Asp-4801) similar to NRPSs. Recognisable PKS domains commencing at the N-terminus of XabB, are an acyl-CoA ligase (AL), acyl carrier protein (ACP1), β-ketoacyl synthase (KS1), and β-ketoacyl reductase (KR), followed by two consecutive ACPs and one KS (Figure 1). The motifs characteristic of these domains are aligned with those from other organisms in Figure 3. The AL domain shows 22-30% identity and 50-60% similarity to prokaryotic and eukaryotic aromatic acid-CoA ligases and long-chain fatty acid-CoA ligases, and contains the conserved adenylation core sequence (SGSSG) and the ATPase motif (TGD). The three ACP domains show up to 39.2% identity and 78.6% similarity to acyl carrier proteins, and all contain a 4'-phosphopantetheinyl binding cofactor box GxDS(I/L) (Hopwood and Sherman, 1990), except that A replaces G in ACP1 (Figure 3). The two KS domains show up to 56.1% identity and 80.8% similarity to β-ketoacyl synthases. Both contain motif GPxxxxxxCSxSL around the active site Cys, and two His residues downstream of the active site Cys, in motifs characteristic of these enzymes (Donadio *et al.*, 1991; Hopwood, 1997; Huang *et al.*, 1998). The KR domain shows up to 27.9% identity and 61.8% similarity to β-ketoacyl reductases, and contains the NAD(P)H binding site GGxGxLG (Scrutton *et al.*, 1990).

At the C-terminus of XabB is an apparent peptide synthetase region linked to the PKS module *via* a peptidyl carrier protein (PCP) domain (Figure 1). The peptide synthetase region shows 31-38% identity and 60-63% similarity with members of the peptide synthetase family. It displays the ordered condensation, adenylation, and PCP domains typical of such multienzymes (Marahiel *et al.*, 1997) followed by an extra condensation domain. The conserved sequences, characteristic of the domains commonly found in peptide synthetases, are compared with those from XabB in Table 2.

In more detail, the full-length amino acid sequence of the *X. albilineans* PKS-NRPS, presented in SEQ ID NO: 2, extends 4801 residues and includes the following sequence signature motifs:

(a) acyl-CoA ligase (AL) motif I extending from about residue 226 to about residue 240, and motif II extending from about residue 486 to about residue 493;

(b) β-ketoacyl synthase 1 (KS1) motif I extending from about residue 897 to about residue 913, motif II extending from about residue 1038 to about residue 1047, and motif III extending from about residue 1080 to about residue 1089;

(c) β-ketoacyl synthase 2 (KS2) motif I extending from about residue 2777 to about residue 2793, motif II extending from about residue 2918 to about residue 2927, and motif III extending from about residue 2955 to about residue 2964;

(d) β-ketoacyl reductase (KR) motif extending from about residue 1812 to about residue 1842;

(e) acyl carrier protein 1 (ACP1) motif extending from about residue 667 to about residue 678;

(f) acyl carrier protein 2 (ACP2) motif extending from about residue 2484 to about residue 2495;

(g) acyl carrier protein 3 (ACP3) motif extending from about residue 2568 to about residue 2579;

(h) adenylation domain (A) motif I extending from about residue 3806 to about residue 3811, motif II extending from about residue 3851 to about residue 3861, motif III extending from about residue 3917 to about residue 3932; motif IV extending from about residue 3967 to about residue 3970, motif V extending from about residue 4063 to about residue 4069, motif VI extending from about residue 4114 to about residue 4128, motif VII extending from about residue 4152 to about residue 4157, motif VIII

extending from about residue 4170 to about residue 4189, motif IX extending from about residue 4239 to about residue 4245, and motif X extending from about residue 4259 to about residue 4264;

(i) peptidyl carrier protein 1 (PCP1) motif extending from about residue 3261 to about residue 3271;

(j) peptidyl carrier protein 2 (PCP2) motif extending from about residue 4306 to about residue 4316;

(k) condensation domain 1 (C1) motif I extending from about residue 3333 to about residue 3342, motif II extending from about residue 3381 to about residue 3389, and motif III extending from about residue 3456 to about residue 3465, motif IV extending from about residue 3495 to about residue 3501, motif V extending from about residue 3606 to about residue 3617, motif VI extending from about residue 3641 to about residue 3647, motif VII extending from about residue 3658 to about residue 3665; and

(l) condensation domain 2 (C2) motif I extending from about residue 4374 to about residue 4383, motif II extending from about residue 4421 to about residue 4429, and motif III extending from about residue 4498 to about residue 4507, motif IV extending from about residue 4538 to about residue 4544, motif V extending from about residue 4649 to about residue 4659, motif VI extending from about residue 4685 to about residue 4691, motif VII extending from about residue 4701 to about residue 4708.

From the above signature motifs, it can be deduced that XabB commences with an AL domain (residues 1-629) followed by an ACP domain (ACP1, residues 630-731). In other PKS systems, an N-terminal AL is involved in activation and incorporation of 3,4-dihydroxycyclohexane carboxylic acid, 3-amino-5-hydroxy benzoic acid (AHBA), or long-chain fatty acid as a starter (Aparicio *et al.*, 1996; Motamedi and Shafiee, 1998; Tang *et al.*, 1998; Duitman *et al.*, 1999). The second module in XabB contains a KS (residues 732-1165), and a KR (residues 1811-1971) upstream of two ACPs (residues 2457-2522, 2544-2613), but lacks any discemable AT domain (Figure 1). The third module contains a KS (residues 2630-3046) followed by a PCP (residues 3221-3307) at the start of the XabB NRPS region.

Four other fused PKS/NRPS systems (Albertini *et al.*, 1995; Gehring *et al.*, 1998; Duitman *et al.*, 1999; Paitan *et al.*, 1999) are known, three of which lack recognisable AT domains (Figure 6). *Yersinia pestis* HMWP1 contains a typical PKS elongation module

(including AT), and an NRPS module with a terminating TE domain. It is the third protein, following an AL (YbtE) and NRPS (HMWP2) in the biosynthetic apparatus for yersiniabactin (Gehring *et al.*, 1998). *B. subtilis* MycA bears the closest resemblance to XabB, showing PKS initiation and elongation modules linked *via* an amino transferase (AMT) domain to the NRPS region. In *B. subtilis* PksK and *M. xanthus* Ta1, the NRPS region precedes the PKS region. Separate AT enzymes encoded elsewhere in the genome may operate in *trans* to load the appropriate acyl groups onto the ACPs in the elongation modules of these PKSs. Candidates are a malonyl-CoA tranascylase gene (*fenF*) located immediately upstream of *mycA* (Duitman *et al.*, 1999), and an acyltransferase gene located 20 kb upstream of *ta1* (Paitan *et al.*, 1999). Accordingly, it is believed that one or more such *trans*-acting AT enzymes may also be involved in connection with the operation of XabB.

From the characteristics of albicidin, and the architecture of the XabB PKS region (Figure 1), the inventor considers that: (i) the AL couples coenzyme A to a shikimate-derived acyl residue in an ATP-dependent reaction, and loads the activated acyl unit onto the 4'-phosphopantetheine prosthetic arm of ACP1; (ii) an acyl group is loaded onto ACP2 or ACP3 by a separate acyltransferase; (iii) the KS1 domain accepts the acyl residue from ACP1 onto a conserved cysteine residue, then transfers it by decarboxylative condensation onto the acyl group tethered to ACP2 or ACP3; (iv) the tethered chain is modified by KR; (v) the assembled polyketide intermediate is translocated *via* KS2 onto the 4-phosphopantetheine prosthetic arm of PCP1, at the start of the NRPS region.

The A domain in the NRPS region of XabB contains ten conserved sequences (A1 to A10, Table 2) identified as AMP, ATP-Mg binding, adenine binding or ATPase sites (Turgay *et al.*, 1992; Marahiel *et al.*, 1997). In other NRPS systems, A domains select and load a particular amino acid, nonproteinogenic amino, hydroxyl or carboxy acid (Marahiel *et al.*, 1997). Substrate specificity is determined at the binding pocket, consisting of a stretch of about 100 amino acid residues between highly conserved motif A4 and A5 (Conti *et al.*, 1997). Sequence alignments for this region reveal some clusters corresponding with the loaded substrate (Stachelhaus *et al.*, 1999). The A domain from XabB falls in a diverse cluster of NRPS modules involved in loading of His, Leu or aromatic amino acids (Phe and Tyr) in other NRPS systems (Figure 7). Based on the architecture of the XabB NPRS region, it can be inferred that the polyketide intermediate

tethered on PCP1 is accepted by C1 and coupled to the amino, hydroxyl, or carboxy acid preloaded by A onto PCP2. The final condensation domain at the C-terminus of XabB is probably involved in peptide-chain termination and cyclisation, as in enniatin, HC-toxin, rapamycin and FK506 systems (Konz and Marahiel, 1999).

2.1.2 Phosphopantetheinyl transferase associated with albicidin biosynthesis

The present invention also provides a gene (*xabA*) from *X. albilineans* encoding a phosphopantetheinyl transferase (PPTase) associated with XabB function. In this regard, XabB contains five carrier protein (ACP/PCP) domains, to which the growing polyketide or polypeptide chain could be covalently tethered. Each functional ACP or PCP domain must have a specific serine side chain phosphopantetheinylated by a dedicated PPTase (Lambalot *et al.*, 1996). The product of *xabA* (XabA) fulfils this function and is required for post-translational activation of synthetases in the albicidin biosynthetic pathway.

The full-length amino acid sequence of this *X. albilineans* PPTase, presented in SEQ ID NO: 83, extends 278 residues and includes the sequence signature motifs for PPTases which are located as follows: (I) motif I spanning from about residue 159 to about residue 167; and (II) motif II spanning from about residue 207 to about residue 218, of SEQ ID NO: 83. The sequence intervening between the two motifs extends from about residue 168 to about residue 206 of SEQ ID NO: 83. These conserved sequence motifs and the intervening sequence are presented for convenience in SEQ ID NO: 89, 93 and 91, respectively.

The deduced *xabA* gene product has 56-62 % overall similarity to EntD proteins for enterobactin biosynthesis and 39-56 % overall similarity to other enzymes in the phosphopantetheinyl transferase superfamily. Like *entD*, *xabA* includes rarely used codons, which may impose post-transcriptional control on the rate of gene product formation (Coderre & Earhart, 1989). Codon optimisation of *xabA* may, therefore, be useful for enhancing the production of XabA.

2.1.3 Methyltransferase associated with albicidin biosynthesis

The invention also provides a gene (*xabC*) from *X. albilineans* encoding a methyltransferase enzyme, more particularly an *O*-methyltransferase enzyme, which is

required for albicidin production and which when expressed above natural levels leads to increased levels and/or functional activities of albicidin antibiotics. The full-length amino acid sequence of this *X. albilineans* methyltransferase, presented in SEQ ID NO: 95, extends 343 residues and includes methyltransferase consensus sequence motifs which are located as follows: (I) motif I spanning from about residue 173 to about residue 180; (II) motif II spanning from about residue 236 to about residue 243; and (III) motif III spanning from about residue 266 to about residue 274, of SEQ ID NO: 95. These conserved sequence motifs are presented for convenience in SEQ ID NO: 99, 101 and 103, respectively.

2.2 Biologically active fragments

The invention also contemplates biological fragments of the above polypeptides of at least 6 and preferably at least 8 amino acids in length, which comprise an activity associated with the domains described above. For example, biologically active fragments may be produced according to any suitable procedure known in the art. For example, a suitable method may include first producing a fragment of a parent polypeptide as described in Section 2.1 and then testing the fragment for the appropriate biological activity. In one embodiment, the fragment is derived from the albicidin PKS-NRPS of the invention and is tested for an activity selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity and condensation activity.

Any assays that detects or preferably measure such activities is contemplated in the practice of the present invention. The biologically active fragment suitably comprises any one or more of the sequence signature motifs described above, or variants thereof. Preferably, the biologically active fragment comprises all said sequence signature motifs, or variants thereof.

In another embodiment, the fragment is derived from the PPTase of the invention and is tested for PPTase activity according to standard assays known to persons of skill in the art. Suitably, the PPTase catalyses the pantetheinylation, more preferably the phosphopantetheinylation, of proteins involved in antibiotic biosynthesis, preferably albicidin biosynthesis. The biologically active fragment preferably comprises the consensus sequence motifs set forth in SEQ ID NO: 89 and 93, or variant thereof and thus,

more preferably comprises the sequence from about residue 159 to about residue 218, of SEQ ID NO: 83.

In yet another embodiment, the fragment is derived from the methyltransferase of the invention and is tested for methyltransferase activity, preferably *O*-methyltransferase activity and more preferably *S*-adenosylmethionine-dependent *O*-methyltransferase activity. Suitably, the methyltransferase catalyses the transfer of one or more methyl groups to an antibiotic precursor, more preferably an albicidin precursor or an intermediate relating to the biosynthesis of albicidins. The biologically active fragment preferably comprises the consensus sequence motifs set forth in SEQ ID NO: 99, 101 and 103, or variant thereof and thus, more preferably comprises the sequence from about residue 173 to about residue 274 of SEQ ID NO: 95 (*i.e.*, SEQ ID NO: 105), or variant of said sequence. In an especially preferred embodiment, the biologically active fragment comprises the sequence from about residue 1 to about residue 277 of SEQ ID NO: 95 (*i.e.*, SEQ ID NO: 107), or variant of said sequence. An exemplary polynucleotide encoding this sequence is cloned in vector pLXABB described *infra*.

Alternatively, biological activity of the fragment is tested by introducing a polynucleotide from which a fragment of a parent polypeptide can be translated into a cell, and detecting one or more of the above activities, which is indicative of said fragment being a biologically active fragment. In one embodiment, such activity can be assayed by introducing into an albicidin deficient *xabB*$^-$ *X. albilineans* mutant (*e.g.*, strain LS157 described herein) a polynucleotide from which a PKS-NRPS-associated fragment can be produced and assaying for antibiotic activity using a microbial plate assay, as for instance described in Example 1.

In another embodiment embodiment, PPTase activity is assayed by introducing into an albicidin deficient *xabA*$^-$ *X. albilineans* mutant (*e.g.*, strain LS156 described herein) a polynucleotide from which a PPTase-associated fragment can be produced and assaying for antibiotic activity using a microbial plate assay, as for instance described in Example 2.

In yet another embodiment, methyltransferase activity is assayed by introducing into an albicidin deficient *xabC*$^-$ *X. albilineans* mutant (*e.g.*, strain LS-JP1 described herein) a polynucleotide from which a methyltransferase-associated fragment can be

produced and assaying for antibiotic activity as for example described herein using a microbial plate assay, as for instance described in Example 3.

2.3 Polypeptide variants

The invention also contemplates polypeptide variants of the polypeptides of the invention wherein said variants have an activity selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity, and methyltransferase activity. Suitable methods of producing polypeptide variants include, for example, producing a modified polypeptide whose sequence is distinguished from a parent polypeptide as described in Section 2.1 or a biologically active fragment thereof by the substitution, deletion and/or addition of at least one amino acid. The modified polypeptide is then tested for one or more of said activities, wherein the presence of that activity indicates that the modified polypeptide is a variant of the parent polypeptide.

In another embodiment, a polypeptide variant is produced by introducing into a cell a polynucleotide from which a modified polypeptide can be translated, and detecting one or more of the activities described above that are associated with the cell, which is indicative of the modified polypeptide being a polypeptide variant.

In general, variants will have at least 60%, more suitably at least 70%, preferably at least 80%, and more preferably at least 90% homology to a polypeptide as for example shown in SEQ ID NO: 4, or a biological fragment thereof. It is preferred that variants display at least 60%, more suitably at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90% and still more preferably at least 95% sequence identity with a parent polypeptide as described in Section 2.1 or a biologically active fragment thereof. In this respect, the window of comparison preferably spans about the full length of the polypeptide or of the biologically active fragment. Suitable variants can be obtained from any secondary metabolite-producing organism, and preferably from an albicidin-producing organism.

Alternatively polypeptide variants according to the invention can be identified either rationally, or *via* established methods of mutagenesis (see, for example, Watson, J.

D. *et al.*, "MOLECULAR BIOLOGY OF THE GENE", Fourth Edition, Benjamin/Cummings, Menlo Park, Calif., 1987). Significantly, a random mutagenesis approach requires no *a priori* information about the gene sequence that is to be mutated. This approach has the advantage that it assesses the desirability of a particular mutant based on its function, and thus does not require an understanding of how or why the resultant mutant protein has adopted a particular conformation. Indeed, the random mutation of target gene sequences has been one approach used to obtain mutant proteins having desired characteristics (Leatherbarrow, R. 1986, *J. Prot. Eng.* **1**: 7-16; Knowles, J. R., 1987, *Science* **236**: 1252-1258; Shaw, W. V., 1987, *Biochem. J.* **246**: 1-17; Gerit, J. A. 1987, *Chem. Rev.* **87**: 1079-1105). Alternatively, where a particular sequence alteration is desired, methods of site-directed mutagenesis can be employed. Thus, such methods may be used to selectively alter only those amino acids of the protein that are believed to be important (Craik, C. S., 1985, *Science* **228**: 291-297; Cronin, *et al.*, 1988, *Biochem.* **27**: 4572-4579; Wilks, *et al.*, 1988, *Science* **242**: 1541-1544).

Variant peptides or polypeptides, resulting from rational or established methods of mutagenesis or from combinatorial chemistries may comprise conservative amino acid substitutions. Exemplary conservative substitutions in a polypeptide or polypeptide fragment according to the invention may be made according to the following table:

TABLE B

Original Residue	*Exemplary Substitutions*
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro

Original Residue	*Exemplary Substitutions*
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile,
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

Substantial changes in function are made by selecting substitutions that are less conservative than those shown in TABLE B. Other replacements would be non-conservative substitutions and relatively fewer of these may be tolerated. Generally, the substitutions which are likely to produce the greatest changes in a polypeptide's properties are those in which (a) a hydrophilic residue (*e.g.*, Ser or Asn) is substituted for, or by, a hydrophobic residue (*e.g.*, Ala, Leu, Ile, Phe or Val); (b) a cysteine or proline is substituted for, or by, any other residue; (c) a residue having an electropositive side chain (*e.g.*, Arg, His or Lys) is substituted for, or by, an electronegative residue (*e.g.*, Glu or Asp) or (d) a residue having a smaller side chain (*e.g.*, Ala, Ser) or no side chain (*e.g.*, Gly) is substituted for, or by, one having a bulky side chain (*e.g.*, Phe or Trp).

2.4 Polypeptide derivatives

A polypeptide can typically tolerate one or more amino acid deletions and insertions in its amino acid sequence without loss or significant loss of a desired activity. Accordingly, the invention also contemplates derivatives of the parent polypeptides of the invention described in Section 2.1 or biologically active fragments thereof or variants of

these, which include amino acid deletions and/or additions, wherein said derivatives comprise one or more activities selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity associated with antibiotic biosynthesis, and preferably with albicidin biosynthesis.

Preferred derivatives of the invention include PKS-NRPS molecules with altered activities in one or more respects and thus produce polyketides other than the albicidin natural product(s) of the XabB. A PKS-NRPS derived from XabB by such alteration includes a modular PKS-NRPS (or its corresponding encoding gene(s)) that retains the scaffolding of the utilised portion encoded by the naturally occurring gene. Not all domains or modules need be altered. On the constant scaffold, at least one enzymatic activity is mutated, deleted, replaced, or inserted so as to alter the activity of the resulting PKS-NRPS relative to the original or parent PKS-NRPS. Alteration results when these activities are deleted or are replaced by a different version of the activity, or simply mutated in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or elongation unit, stereochemistry, chain length or cyclisation, and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a deleted activity is replaced, the origin of the replacement activity may come from a corresponding activity in a different naturally occurring PKS or PKS-NRPS or from a different region of the albicidin PKS-NRPS. Any or all PKS/NRPS genes may be included in the derivative or portions of any of these may be included, but the scaffolding of the albicidin PKS-NRPS protein is preferably retained in whatever derivative is constructed.

Thus, a PKS-NRPS derived from the albicidin PKS-NRPS includes a PKS-NRPS that contains the scaffolding of all or a portion of XabB. The derived PKS-NRPS also contains at least two elongation modules that are functional and preferably at least three elongation modules. The derived PKS-NRPS also contains mutations, deletions, insertions, or replacements of one or more of the activities of the functional domains or modules of XabB so that the nature of the resulting polyketide is altered. Exemplary embodiments include those wherein a KS or ACP domain has been deleted or replaced by a version of the activity from a different PKS/NRPS or from another location within XabB. Also

contemplated are derivatives where at least one non-condensation cycle enzymatic activity (KR, KR, or A) has been deleted or added or wherein any of these activities has been mutated so as to change the structure of the polyketide synthesised by the PKS.

Other derivatives contemplated by the present invention include fusion of the polypeptides, fragments and polypeptide variants of the invention with other polypeptides or proteins. For example, it will be appreciated that said polypeptides, fragments or variants may be incorporated into larger polypeptides, and that such larger polypeptides may also be expected to have one or more of the activities mentioned above. The polypeptides, fragments or variants of the invention may be fused to a further protein, for example, which is not derived from the original host. The further protein may assist in the purification of the fusion protein. For instance, a polyhistidine tag or a maltose binding protein may be used in this respect as described in more detail below. Other possible fusion proteins are those which produce an immunomodulatory response. Particular examples of such proteins include Protein A or glutathione S-transferase (GST).

Other derivatives contemplated by the invention include, but are not limited to, modification to side chains, incorporation of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the polypeptides, fragments and variants of the invention. Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by acylation with acetic anhydride; acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; amidination with methylacetimidate; carbamoylation of amino groups with cyanate; pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$; reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; and trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS). The carboxyl group may be modified by carbodiimide activation via O-acylisourea formation followed by subsequent derivatisation, by way of example, to a corresponding amide. The guanidine group of arginine residues may be modified by formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal. Sulphydryl groups may be modified by methods such as performic acid oxidation to cysteic acid; formation of mercurial derivatives using 4-chloromercuriphenylsulphonic acid, 4-chloromercuribenzoate; 2-chloromercuri-4-

nitrophenol, phenylmercury chloride, and other mercurials; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; carboxymethylation with iodoacetic acid or iodoacetamide; and carbamoylation with cyanate at alkaline pH. Tryptophan residues may be modified, for example, by alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphonyl halides or by oxidation with N-bromosuccinimide. Tyrosine residues may be modified by nitration with tetranitromethane to form a 3-nitrotyrosine derivative. The imidazole ring of a histidine residue may be modified by N-carbethoxylation with diethylpyrocarbonate or by alkylation with iodoacetic acid derivatives.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include but are not limited to, use of 4-amino butyric acid, 6-aminohexanoic acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 4-amino-3-hydroxy-6-methylheptanoic acid, t-butylglycine, norleucine, norvaline, phenylglycine, ornithine, sarcosine, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids contemplated by the present invention is shown in TABLE C.

TABLE C

Non-conventional amino acid	*Non-conventional amino acid*
α-aminobutyric acid	L-N-methylalanine
α-amino-α-methylbutyrate	L-N-methylarginine
aminocyclopropane-carboxylate	L-N-methylasparagine
aminoisobutyric acid	L-N-methylaspartic acid
aminonorbornyl-carboxylate	L-N-methylcysteine
cyclohexylalanine	L-N-methylglutamine
cyclopentylalanine	L-N-methylglutamic acid
L-N-methylisoleucine	L-N-methylhistidine
D-alanine	L-N-methylleucine
D-arginine	L-N-methyllysine
D-aspartic acid	L-N-methylmethionine

Non-conventional amino acid	*Non-conventional amino acid*
D-cysteine	L-N-methylnorleucine
D-glutamate	L-N-methylnorvaline
D-glutamic acid	L-N-methylornithine
D-histidine	L-N-methylphenylalanine
D-isoleucine	L-N-methylproline
D-leucine	L-N-medlylserine
D-lysine	L-N-methylthreonine
D-methionine	L-N-methyltryptophan
D-ornithine	L-N-methyltyrosine
D-phenylalanine	L-N-methylvaline
D-proline	L-N-methylethylglycine
D-serine	L-N-methyl-t-butylglycine
D-threonine	L-norleucine
D-tryptophan	L-norvaline
D-tyrosine	α-methyl-aminoisobutyrate
D-valine	α-methyl-γ-aminobutyrate
D-α-methylalanine	α-methylcyclohexylalanine
D-α-methylarginine	α-methylcylcopentylalanine
D-α-methylasparagine	α-methyl-α-napthylalanine
D-α-methylaspartate	α-methylpenicillamine
D-α-methylcysteine	N-(4-aminobutyl)glycine
D-α-methylglutamine	N-(2-aminoethyl)glycine
D-α-methylhistidine	N-(3-aminopropyl)glycine
D-α-methylisoleucine	N-amino-α-methylbutyrate
D-α-methylleucine	α-napthylalanine

Non-conventional amino acid	*Non-conventional amino acid*
D-α-methyllysine	N-benzylglycine
D-α-methylmethionine	N-(2-carbamylediyl)glycine
D-α-methylornithiine	N-(carbamylmethyl)glycine
D-α-methylphenylalanine	N-(2-carboxyethyl)glycine
D-α-methylproline	N-(carboxymethyl)glycine
D-α-methylserine	N-cyclobutylglycine
D-α-methylthreonine	N-cycloheptylglycine
D-α-methyltryptophan	N-cyclohexylglycine
D-α-methyltyrosine	N-cyclodecylglycine
L-α-methylleucine	L-α-methyllysine
L-α-methylmethionine	L-α-methylnorleucine
L-α-methylnorvatine	L-α-methylornithine
L-α-methylphenylalanine	L-α-methylproline
L-α-methylserine	L-α-methylthreonine
L-α-methyltryptophan	L-α-methyltyrosine
L-α-methylvaline	L-N-methylhomophenylalanine
N-(N-(2,2-diphenylethyl carbamylmethyl)glycine	N-(N-(3,3-diphenylpropyl carbamylmethyl)glycine
1-carboxy-1-(2,2-diphenyl-ethyl amino)cyclopropane	

Also contemplated is the use of crosslinkers, for example, to stabilise 3D conformations of the polypeptides, fragments or variants of the invention, using homo-bifunctional cross linkers such as bifunctional imido esters having $(CH_2)_n$ spacer groups with n = 1 to n = 6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety or

carbodiimide. In addition, peptides can be conformationally constrained, for example, by introduction of double bonds between C_α and C_β atoms of amino acids, by incorporation of C_α and N_α-methylamino acids, and by formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini between two side chains or between a side chain and the N or C terminus of the peptides or analogues. For example, reference may be made to: Marlowe (1993, *Biorganic & Medicinal Chemistry Letters* **3**: 437-44) who describes peptide cyclisation on TFA resin using trimethylsilyl (TMSE) ester as an orthogonal protecting group; Pallin and Tam (1995, *J. Chem. Soc. Chem. Comm.* 2021-2022) who describe the cyclisation of unprotected peptides in aqueous solution by oxime formation; Algin *et al* (1994, *Tetrahedron Letters* **35**: 9633-9636) who disclose solid-phase synthesis of head-to-tail cyclic peptides *via* lysine side-chain anchoring; Kates *et al* (1993, *Tetrahedron Letters* **34**: 1549-1552) who describe the production of head-to-tail cyclic peptides by three-dimensional solid phase strategy; Tumelty *et al* (1994, *J. Chem. Soc. Chem. Comm.* 1067-1068) who describe the synthesis of cyclic peptides from an immobilised activated intermediate, wherein activation of the immobilised peptide is carried out with *N*-protecting group intact and subsequent removal leading to cyclisation; McMurray *et al* (1994, *Peptide Research* **7**: 195-206) who disclose head-to-tail cyclisation of peptides attached to insoluble supports by means of the side chains of aspartic and glutamic acid; Hruby *et al* (1994, *Reactive Polymers* **22**: 231-241) who teach an alternate method for cyclising peptides *via* solid supports; and Schmidt and Langer (1997, *J. Peptide Res.* **49**: 67-73) who disclose a method for synthesising cyclotetrapeptides and cyclopentapeptides. The foregoing methods may be used to produce conformationally constrained polypeptides that comprise one or more activities selected form the group consisting of acyl-CoA ligase activity, β-ketoacyl synthase activity, β-ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity associated with the production of polyketides and particularly albicidins or analogues thereof.

The invention also contemplates polypeptides, fragments or variants of the invention that have been modified using ordinary molecular biological techniques so as to improve their resistance to proteolytic degradation or to optimise solubility properties or to render them more suitable as an immunogenic agent.

3. *Polynucleotides of the invention*

3.1 Polynucleotides encoding polypeptides of the invention

3.1.1 Albicidin synthetase-encoding polynucleotides

The invention further provides a polynucleotide that encodes a PKS-NRPS polypeptide of the invention, or biologically active fragment thereof, or variant or derivative of these as defined above. In one embodiment, the polynucleotide comprises the entire sequence of nucleotides set forth in SEQ ID NO: 1. SEQ ID NO: 1 corresponds to a 16511-bp *X. albilineans xabB* cistron. SEQ ID NO: 3, defines the full-length coding sequence of *xabB* and encodes various sequence signature motifs at the following nucleotide positions:

(a) acyl-CoA ligase (AL) motif I from about nucleotide 676 to about nucleotide 720, and motif II from about nucleotide 1456 to about nucleotide 1477;

(b) β-ketoacyl synthase 1 (KS1) motif I from about nucleotide 2689 to about nucleotide 2739, motif II from about nucleotide 3112 to about nucleotide 3141, and motif III from about nucleotide 3238 to about nucleotide 3267;

(c) β-ketoacyl synthase 2 (KS2) motif I from about nucleotide 8329 to about nucleotide 8379, motif II from about nucleotide 8752 to about nucleotide 8781, and motif III from about nucleotide 8863 to about nucleotide 8892;

(d) β-ketoacyl reductase (KR) motif from about nucleotide 5434 to about nucleotide 5526;

(e) acyl carrier protein 1 (ACP1) motif from about nucleotide 1999 to about nucleotide 2034;

(f) acyl carrier protein 2 (ACP2) motif from about nucleotide 7450 to about nucleotide 7485;

(g) acyl carrier protein 3 (ACP3) motif from about nucleotide 7702 to about nucleotide 7735;

(h) adenylation domain (A) motif I from about nucleotide 11416 to about nucleotide 11433, motif II from about nucleotide 11551 to about nucleotide 11583, motif III from about nucleotide 11749 to about nucleotide 11796; motif IV from about nucleotide 11899 to about nucleotide 11910, motif V from about nucleotide 12187 to about nucleotide 12207, motif VI from about nucleotide 12340 to about nucleotide 12384,

motif VII from about nucleotide 12454 to about nucleotide 12471, motif VIII from about nucleotide 12508 to about nucleotide 12567, motif IX from about nucleotide 12715 to about nucleotide 12735, and motif X from about nucleotide 12775 to about nucleotide 12792;

(i) peptidyl carrier protein 1 (PCP1) motif from about nucleotide 9781 to about nucleotide 9813;

(j) peptidyl carrier protein 2 (PCP2) motif from about nucleotide 12916 to about nucleotide 12948;

(k) condensation domain 1 (C1) motif I from about nucleotide 9997 to about nucleotide 10026, motif II from about nucleotide 10141 to about nucleotide 10167, and motif III from about nucleotide 10366 to about nucleotide 10395, motif IV from about nucleotide 10483 to about nucleotide 10503, motif V from about nucleotide 10816 to about nucleotide 10851, motif VI from about nucleotide 10921 to about nucleotide 10941, motif VII from about nucleotide 10972 to about nucleotide 10995; and

(l) condensation domain 2 (C2) motif I from about nucleotide 13120 to about nucleotide 13149, motif II from about nucleotide 13261 to about nucleotide 13287, and motif III from about nucleotide 13492 to about nucleotide 13521, motif IV from about nucleotide 13612 to about nucleotide 13632, motif V from about nucleotide 13945 to about nucleotide 13977, motif VI from about nucleotide 14053 to about nucleotide 14073, motif VII from about nucleotide 14101 to about nucleotide 14124.

Those of skill in the art will recognise that, due to the degenerate nature of the genetic code, a variety of polynucleotides differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native polynucleotide sequence encoding the PKS-NRPS of *X. albilineans* is shown herein merely to illustrate a preferred embodiment of the invention, and the invention includes polynucleotides of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention.

3.1.2 PPTase-encoding polynucleotides

The invention further provides a polynucleotide that encodes a PPTase polypeptide of the invention, or biologically active fragment thereof, or variant or derivative of these as defined above. In one embodiment, the polynucleotide comprises the

entire sequence of nucleotides set forth in SEQ ID NO: 82. SEQ ID NO: 82 corresponds to a 1200-bp *X. albilineans xabA* cistron. This sequence encodes a PPTase catalytic domain from about nucleotide 475 to about nucleotide 654. This domain comprises two conserved PPTase sequence motifs: (I) motif I encoded by a nucleotide sequence from about nucleotide 475 to about nucleotide 501; and (II) motif II encoded by a nucleotide sequence from about nucleotide 619 to about nucleotide 654, of SEQ ID NO: 82. The intervening amino acid sequence, linking motifs 1 and II, is encoded by a nucleotide sequence from about nucleotide 502 to about nucleotide 618 of SEQ ID NO: 82. The said nucleotide sequences are presented for convenience in SEQ ID NO: 86, 88, 92 and 90, respectively. Suitably, the polynucleotide comprises the sequence set forth in SEQ ID NO: 84, which defines the full-length coding sequence of *xabA*. Alternatively, the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 86, which encodes the PPTase catalytic domain.

3.1.3 Methyltransferase-encoding polynucleotides

The invention further provides a polynucleotide that encodes a methyltransferase polypeptide of the invention, or biologically active fragment thereof, or variant or derivative of these as defined above. In one embodiment, the polynucleotide comprises the entire sequence of nucleotides set forth in SEQ ID NO: 94. SEQ ID NO: 94 corresponds to a 1515-bp *X. albilineans xabC* cistron. This sequence encodes three conserved methyltransferase sequence motifs: (I) motif I encoded by a nucleotide sequence from about nucleotide 565 to about nucleotide 585; (II) motif II encoded by a nucleotide sequence from about nucleotide 741 to about nucleotide 774; and (III) motif III encoded by a nucleotide sequence from about nucleotide 841 to about nucleotide 867, or SEQ ID NO: 94. The said nucleotide sequences are presented for convenience in SEQ ID NO: 98, 100 and 102, respectively. Suitably, the polynucleotide comprises the sequence set forth in SEQ ID NO: 96, which defines the full-length coding sequence of *xabC*. Alternatively, the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 104 or 106, which encode biologically active fragments as described in Section 2.2.

3.2 Polynucleotide variants

In general, polynucleotide variants according to the invention comprise regions that show at least 60%, more suitably at least 70%, preferably at least 80%, and more preferably at least 90% sequence identity over a reference polynucleotide sequence of identical size ("*comparison window*") or when compared to an aligned sequence in which the alignment is performed by a computer homology program known in the art. What constitutes suitable variants may be determined by conventional techniques. For example, a polynucleotide comprising at least one sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102 and 104 can be altered using any suitable method including conventional recombinant techniques and mutagenesis methods such as random mutagenesis (*e.g.*, transposon mutagenesis), oligonucleotide-mediated (or site-directed) mutagenesis, PCR mutagenesis and cassette mutagenesis of an earlier prepared variant or non-variant version of an isolated polynucleotide of the invention.

Alternatively, polynucleotide sequences variants encoding heterologous PKS/NRPS enzymes for producing PKS-NRPS variants of the invention may be obtained from other secondary metabolite- or polyketide-producing organisms. For example, such variants may be prepared according to the following procedure:

(a) creating primers which are optionally degenerate wherein each comprises a portion of a reference polynucleotide encoding a reference polypeptide or fragment of the invention, preferably encoding at least one sequence selected from the group consisting of SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 89, 91, 93, 95, 99, 101, 103, 105 and 107;

(b) obtaining a nucleic acid extract from a secondary metabolite-producing organism, which is preferably a bacterium, more preferably from a species of the family *Pseudomonadaceae*, more preferably from a *Xanthomonas* species; and

(c) using said primers to amplify, *via* nucleic acid amplification techniques, at least one amplification product from said nucleic acid extract, wherein said amplification product corresponds to a polynucleotide variant.

Suitable nucleic acid amplification techniques are well known to the skilled addressee, and include polymerase chain reaction (PCR) as for example described in Ausubel *et al.* (*supra*); strand displacement amplification (SDA) as for example described in U.S. Patent No 5,422,252; rolling circle replication (RCR) as for example described in Liu *et al.*, (1996, *J. Am. Chem. Soc.* **118**:1587-1594 and International application WO 92/01813) and Lizardi *et al.*, (International Application WO 97/19193); nucleic acid sequence-based amplification (NASBA) as for example described by Sooknanan *et al.*, (1994, *Biotechniques* **17**:1077-1080); and Q-β replicase amplification as for example described by Tyagi *et al.*, (1996, *Proc. Natl. Acad. Sci. USA* **93**: 5395-5400).

Typically, polynucleotide variants that are substantially complementary to a reference polynucleotide are identified by blotting techniques that include a step whereby nucleic acids are immobilised on a matrix (preferably a synthetic membrane such as nitrocellulose), followed by a hybridisation step, and a detection step. Southern blotting is used to identify a complementary DNA sequence; northern blotting is used to identify a complementary RNA sequence. Dot blotting and slot blotting can be used to identify complementary DNA/DNA, DNA/RNA or RNA/RNA polynucleotide sequences. Such techniques are well known by those skilled in the art, and have been described in Ausubel *et al.* (1994-1998, *supra*) at pages 2.9.1 through 2.9.20.

According to such methods, Southern blotting involves separating DNA molecules according to size by gel electrophoresis, transferring the size-separated DNA to a synthetic membrane, and hybridising the membrane-bound DNA to a complementary nucleotide sequence labelled radioactively, enzymatically or fluorochromatically. In dot blotting and slot blotting, DNA samples are directly applied to a synthetic membrane prior to hybridisation as above. An alternative blotting step is used when identifying complementary polynucleotides in a cDNA or genomic DNA library, such as through the process of plaque or colony hybridisation. A typical example of this procedure is described in Sambrook *et al.* ("Molecular Cloning. A Laboratory Manual", Cold Spring Harbour Press, 1989) Chapters 8-12.

Typically, the following general procedure can be used to determine hybridisation conditions. Polynucleotides are blotted/transferred to a synthetic membrane, as described above. A reference polynucleotide such as a polynucleotide of the invention is labelled as

described above, and the ability of this labelled polynucleotide to hybridise with an immobilised polynucleotide is analysed. A skilled addressee will recognise that a number of factors influence hybridisation. The specific activity of radioactively labelled polynucleotide sequence should typically be greater than or equal to about 10^8 dpm/mg to provide a detectable signal. A radiolabelled nucleotide sequence of specific activity 10^8 to 10^9 dpm/mg can detect approximately 0.5 pg of DNA. It is well known in the art that sufficient DNA must be immobilised on the membrane to permit detection. It is desirable to have excess immobilised DNA, usually 10 μg. Adding an inert polymer such as 10% (w/v) dextran sulfate (MW 500,000) or polyethylene glycol 6000 during hybridisation can also increase the sensitivity of hybridisation (see Ausubel *supra* at 2.10.10).

To achieve meaningful results from hybridisation between a polynucleotide immobilised on a membrane and a labelled polynucleotide, a sufficient amount of the labelled polynucleotide must be hybridised to the immobilised polynucleotide following washing. Washing ensures that the labelled polynucleotide is hybridised only to the immobilised polynucleotide with a desired degree of complementarity to the labelled polynucleotide. It will be understood that polynucleotide variants according to the invention will hybridise to a reference polynucleotide under at least low stringency conditions. Reference herein to low stringency conditions include and encompass from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridisation at 42° C, and at least about 1 M to at least about 2 M salt for washing at 42° C. Low stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridisation at 65° C, and (i) 2xSSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS for washing at room temperature.

Suitably, the polynucleotide variants hybridise to a reference polynucleotide under at least medium stringency conditions. Medium stringency conditions include and encompass from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridisation at 42° C, and at least about 0.1 M to at least about 0.2 M salt for washing at 55° C. Medium stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridisation at 65° C, and (i) 2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS for washing at 60-65° C.

Preferably, the polynucleotide variants hybridise to a reference polynucleotide under high stringency conditions. High stringency conditions include and encompass from at least about 31% v/v to at least about 50% v/v formamide and from about 0.01 M to about 0.15 M salt for hybridisation at 42° C, and about 0.01 M to about 0.02 M salt for washing at 55° C. High stringency conditions also may include 1% BSA, 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridisation at 65° C, and (i) 0.2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 1% SDS for washing at a temperature in excess of 65° C.

Other stringent conditions are well known in the art. A skilled addressee will recognise that various factors can be manipulated to optimise the specificity of the hybridisation. Optimisation of the stringency of the final washes can serve to ensure a high degree of hybridisation. For detailed examples, see Ausubel *et al., supra* at pages 2.10.1 to 2.10.16 and Sambrook *et al.* (1989, *supra*) at sections 1.101 to 1.104.

While stringent washes are typically carried out at temperatures from about 42° C to 68° C, one skilled in the art will appreciate that other temperatures may be suitable for stringent conditions. Maximum hybridisation rate typically occurs at about 20° C to 25° C below the T_m for formation of a DNA-DNA hybrid. It is well known in the art that the T_m is the melting temperature, or temperature at which two complementary polynucleotide sequences dissociate. Methods for estimating T_m are well known in the art (see Ausubel *et al., supra* at page 2.10.8).

In general, the T_m of a perfectly matched duplex of DNA may be predicted as an approximation by the formula:

$$T_m = 81.5 + 16.6 (\log_{10} M) + 0.41 (\%G+C) - 0.63 (\% \text{ formamide}) - (600/\text{length})$$

wherein: M is the concentration of Na^+, preferably in the range of 0.01 molar to 0.4 molar; %G+C is the sum of guanosine and cytosine bases as a percentage of the total number of bases, within the range between 30% and 75% G+C; % formamide is the percent formamide concentration by volume; length is the number of base pairs in the DNA duplex.

The T_m of a duplex DNA decreases by approximately 1° C with every increase of 1% in the number of randomly mismatched base pairs. Washing is generally carried out at T_m – 15° C for high stringency, or T_m – 30° C for moderate stringency.

In a preferred hybridisation procedure, a membrane (*e.g.*, a nitrocellulose membrane or a nylon membrane) containing immobilised DNA is hybridised overnight at 42° C in a hybridisation buffer (50% deionised formamide, 5xSSC, 5x Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrollidone and 0.1% bovine serum albumin), 0.1% SDS and 200 mg/mL denatured salmon sperm DNA) containing labelled probe. The membrane is then subjected to two sequential medium stringency washes (*i.e.*, 2xSSC, 0.1% SDS for 15 min at 45° C, followed by 2xSSC, 0.1% SDS for 15 min at 50° C), followed by two sequential higher stringency washes (*i.e.*, 0.2xSSC, 0.1% SDS for 12 min at 55° C followed by 0.2xSSC and 0.1%SDS solution for 12 min at 65-68° C.

Methods for detecting a labelled polynucleotide hybridised to an immobilised polynucleotide are well known to practitioners in the art. Such methods include autoradiography, phosphorimaging, and chemiluminescent, fluorescent and colorimetric detection.

4. Expression vectors

The present invention further provides expression vectors designed for genetic transformation of cells, preferably prokaryotic cells, comprising a polynucleotide, fragment or variant according to the invention operably linked to a regulatory polynucleotide. An expression vector is typically a nucleic acid that can be introduced into a host cell or cell-free transcription and translation system. An expression vector can be maintained permanently or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm.

The various components of an expression vector can vary widely, depending on the intended use of the vector and especially the host cell(s) in which the vector is intended to replicate or drive expression. For example, the regulatory polynucleotide, which is used to control expression of a polynucleotide of the invention, will generally be appropriate for the host cell used for expression. Numerous types of appropriate expression vectors and suitable regulatory sequences are known in the art for a variety of host cells. Typically, the

regulatory polynucleotide includes, but is not limited to, promoter sequences, leader or signal sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and termination sequences, and enhancer or activator sequences. Constitutive or inducible promoters as known in the art are contemplated by the invention. The promoters may be either naturally occurring promoters, or hybrid promoters that combine elements of more than one promoter.

In a preferred embodiment, the expression vector is operable in a Gram-negative prokaryotic cell. A variety of prokaryotic expression vectors, which may be used as a basis for constructing the expression vector of the invention. These include but are not limited to a chromosomal vector (*e.g.*, a bacteriophage such as bacteriophage λ), an extrachromosomal vector (*e.g.*, a plasmid or a cosmid expression vector). The expression vector will also typically contain an origin of replication, which allows autonomous replication of the vector, and one or more selectable marker genes that allow phenotypic selection of the transformed cells.

The expression vector may also include a fusion partner (typically provided by the expression vector) so that a recombinant polypeptide is expressed as a fusion polypeptide with said fusion partner. The main advantage of fusion partners is that they assist identification and/or purification of said fusion polypeptide. In order to express said fusion polypeptide, it is necessary to ligate a polynucleotide according to the invention into the expression vector so that the translational reading frames of the fusion partner and the polynucleotide coincide. Well known examples of fusion partners include, but are not limited to, glutathione-S-transferase (GST), Fc potion of human IgG, maltose binding protein (MBP) and hexahistidine (HIS_6), which are particularly useful for isolation of the fusion polypeptide by affinity chromatography. For the purposes of fusion polypeptide purification by affinity chromatography, relevant matrices for affinity chromatography are glutathione-, amylose-, and nickel- or cobalt-conjugated resins respectively. Many such matrices are available in "kit" form, such as the QIAexpress™ system (Qiagen) useful with (HIS_6) fusion partners and the Pharmacia GST purification system. In a preferred embodiment, the recombinant polynucleotide is expressed in the commercial vector pFLAG as described more fully hereinafter. Another fusion partner well known in the art is green fluorescent protein (GFP). This fusion partner serves as a fluorescent "tag" which allows the fusion polypeptide of the invention to be identified by fluorescence microscopy

or by flow cytometry. The GFP tag is useful when assessing subcellular localisation of the fusion polypeptide of the invention, or for isolating cells which express the fusion polypeptide of the invention. Flow cytometric methods such as fluorescence activated cell sorting (FACS) are particularly useful in this latter application. Preferably, the fusion partners also have protease cleavage sites, such as for Factor X_a or Thrombin, which allow the relevant protease to partially digest the fusion polypeptide of the invention and thereby liberate the recombinant polypeptide of the invention therefrom. The liberated polypeptide can then be isolated from the fusion partner by subsequent chromatographic separation. Fusion partners according to the invention also include within their scope "epitope tags", which are usually short peptide sequences for which a specific antibody is available. Well known examples of epitope tags for which specific monoclonal antibodies are readily available include c-Myc, influenza virus, haemagglutinin and FLAG tags.

Preferred host cells for purposes of selecting vector components for expression vectors of the present invention include fungal host cells such as yeast and prokaryotic host cells such as *E. coli* and *X. albilineans*, but mammalian cell cultures can also be used. In hosts such as yeasts, plants, or mammalian cells that ordinarily do not produce modular polyketide synthase enzymes, it may be necessary to provide, also typically by recombinant means, suitable holo-ACP synthases to convert the recombinantly produced PKS to functionality.

The expression vector may be used to transform the desired host cell to produce a recombinant host cell for producing *inter alia* a recombinant polypeptide or polyketides, particularly albicidins or analogues thereof, as described hereinafter.

5. *Methods of preparing the polypeptides of the invention*

Polypeptides of the inventions, including the full-length parent polypeptides described in Section 2.1, or their biologically active fragments comprising, for example one or more domains (or fragments of such domains), or variants or derivatives of these, may be prepared by any suitable procedure known to those of skill in the art. For example, the polypeptides may be prepared by a procedure including the steps of: -

(a) preparing a recombinant polynucleotide comprising a nucleotide sequence encoding a polypeptide comprising the sequence set forth in any one of SEQ ID NO: 4 or a fragment thereof comprising at least one sequence selected from the group

consisting of SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 89, 91, 93, 95, 99, 101, 103, 105 and 107, or variant or derivative of these, which nucleotide sequence is operably linked to a regulatory polynucleotide;

(b) introducing the recombinant polynucleotide into a suitable host cell;

(c) culturing the host cell to express recombinant polypeptide from said recombinant polynucleotide; and

(d) isolating the recombinant polypeptide.

Suitably, said nucleotide sequence comprises at least one sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102 and 104.

The recombinant polynucleotide is preferably in the form of an expression vector, which includes a self-replicating extra-chromosomal vector such as a plasmid, or a vector that integrates into a host genome, as for example described above in Section 4. The step of introducing the recombinant polynucleotide into the host cell may be effected by any suitable means including transfection, and transformation, the choice of which will be dependent on the host cell employed. Such methods are well known to those of skill in the art.

Recombinant polypeptides of the invention may be produced by culturing a host cell transformed with an expression vector containing nucleic acid encoding a polypeptide, biologically active fragment, variant or derivative according to the invention. The conditions appropriate for protein expression will vary with the choice of expression vector and the host cell. This is easily ascertained by one skilled in the art through routine experimentation.

Suitable host cells for expression may be prokaryotic or eukaryotic. One preferred host cell for expression of a polypeptide according to the invention is a bacterium. The bacterium used may be *Escherichia coli*. Alternatively, the host cell may be an insect cell such as, for example, *SF9* cells that may be utilised with a baculovirus expression system.

The recombinant protein may be conveniently prepared by a person skilled in the art using standard protocols as for example described in Sambrook, *et al.*, MOLECULAR CLONING. A LABORATORY MANUAL (Cold Spring Harbor Press, 1989), in particular Sections 16 and 17; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (John Wiley & Sons, Inc. 1994-1998), in particular Chapters 10 and 16; and Coligan *et al.*, CURRENT PROTOCOLS IN PROTEIN SCIENCE (John Wiley & Sons, Inc. 1995-1997), in particular Chapters 1, 5 and 6.

Alternatively, the polypeptide, fragments, variants or derivatives of the invention may be synthesised using solution synthesis or solid phase synthesis as described, for example, in Chapter 9 of Atherton and Shephard (*supra*) and in Roberge *et al* (1995, *Science* **269**: 202).

6. *Antigen-binding molecules*

The invention also contemplates antigen-binding molecules that bind specifically to the aforementioned polypeptides, fragments, variants and derivatives. Preferably, an antigen-binding molecule according to the invention is immuno-interactive with any one or more of the amino acid sequences set forth in SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 89, 91, 93, 95, 99, 101, 103, 105 and 107, or variants thereof.

For example, the antigen-binding molecules may comprise whole polyclonal antibodies. Such antibodies may be prepared, for example, by injecting a polypeptide, fragment, variant or derivative of the invention into a production species, which may include mice or rabbits, to obtain polyclonal antisera. Methods of producing polyclonal antibodies are well known to those skilled in the art. Exemplary protocols which may be used are described for example in Coligan *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY, (John Wiley & Sons, Inc, 1991), and Ausubel *et al.*, (1994-1998, *supra*), in particular Section III of Chapter 11.

In lieu of the polyclonal antisera obtained in the production species, monoclonal antibodies may be produced using the standard method as described, for example, by Köhler and Milstein (1975, *Nature* **256**, 495-497), or by more recent modifications thereof as described, for example, in Coligan *et al.*, (1991, *supra*) by immortalising spleen or other

antibody producing cells derived from a production species which has been inoculated with one or more of the polypeptides, fragments, variants or derivatives of the invention.

The invention also contemplates as antigen-binding molecules Fv, Fab, Fab' and $F(ab')_2$ immunoglobulin fragments. Alternatively, the antigen-binding molecule may be in the form of a synthetic stabilised Fv (scFv) fragment, a disulphide stabilised Fv (dsFv) fragment, a diabody (dAb), a minibody and the like, or may comprise non-immunoglobulin derived, protein frameworks. The antigen-binding molecules of the invention may be used for affinity chromatography in isolating a natural or recombinant polypeptide or biologically active fragment of the invention. For example reference may be made to immunoaffinity chromatographic procedures described in Chapter 9.5 of Coligan *et al.*, (1995-1997, *supra*). The antigen-binding molecules can be used to screen expression libraries for variant polypeptides of the invention as described herein. They can also be used to detect polypeptides, fragments, variants and derivatives of the invention as described hereinafter.

7. *Identification of modulators*

The invention also contemplates a method of screening for an agent that modulates the expression of a gene selected from *xabB*, *xabA*, or *xabC*, or a gene belonging to the same regulatory or biosynthetic pathway as *xabB*, *xabA*, or *xabC*, or a variant of that gene, or that modulates the level and/or functional activity of an expression product of that gene or its variant. The method comprises contacting a preparation comprising said expression product (*e.g.*, polypeptide or transcript), or a biologically active fragment thereof, or variant or derivative of these, or a genetic sequence that modulates the expression of said gene (*e.g.*, the natural promoter relating to said gene, *e.g.*, the *xabB* promoter, comprising the sequence set forth in SEQ ID NO: 81 or complement thereof), with a test agent, and detecting a change in the level and/or functional activity of said polypeptide or biologically active fragment thereof, or variant or derivative, or of a product expressed from said genetic sequence.

Modulators contemplated by the present invention includes agonists and antagonists of gene expression include antisense molecules, ribozymes and co-suppression molecules, as for example described in Section 2. Agonists include molecules which increase promoter activity or interfere with negative mechanisms. Agonists of a gene

include molecules which overcome any negative regulatory mechanism. Antagonists of polypeptides encoded by a gene of interest include antibodies and inhibitor peptide fragments.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 Dalton. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including, but not limited to: peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogues or combinations thereof.

Small (non-peptide) molecule modulators of a polypeptide according to the invention, or portion, or domain or module thereof are particularly preferred. In this regard, small organic molecules typically have the ability to gain entry into an appropriate cell and affect the expression of a gene (*e.g.*, by interacting with the regulatory region or transcription factors involved in gene expression); or affect the activity of a gene by inhibiting or enhancing the binding of accessory molecules.

Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogues. Screening may also be directed to known pharmacologically active compounds and chemical analogues thereof.

Screening for modulatory agents according to the invention can be achieved by any suitable method. For example, the method may include contacting a cell comprising a polynucleotide corresponding to a gene as defined above, with an agent suspected of having said modulatory activity and screening for the modulation of the level and/or

functional activity of a protein encoded by said polynucleotide, or the modulation of the level of an expression product encoded by the polynucleotide, or the modulation of the activity or expression of a downstream cellular target of said protein or said expression product. Detecting such modulation can be achieved utilising techniques including, but not restricted to, ELISA, cell-based ELISA, filter-binding ELISA, inhibition ELISA, Western blots, immunoprecipitation, slot or dot blot assays, immunostaining, RIA, scintillation proximity assays, fluorescent immunoassays using antigen-binding molecule conjugates or antigen conjugates of fluorescent substances such as fluorescein or rhodamine, Ouchterlony double diffusion analysis, immunoassays employing an avidin-biotin or a streptavidin-biotin detection system, and nucleic acid detection assays including reverse transcriptase polymerase chain reaction (RT-PCR).

It will be understood that a polynucleotide from which a target molecule of interest is regulated or expressed may be naturally occurring in the cell which is the subject of testing or it may have been introduced into the host cell for the purpose of testing. Further, the naturally-occurring or introduced sequence may be constitutively expressed – thereby providing a model useful in screening for agents which down-regulate expression of an encoded product of the sequence wherein said down regulation can be at the nucleic acid or expression product level – or may require activation – thereby providing a model useful in screening for agents that up-regulate expression of an encoded product of the sequence. Further, to the extent that a polynucleotide is introduced into a cell, that polynucleotide may comprise the entire coding sequence which codes for a target polypeptide or it may comprise a portion of that coding sequence (*e.g.* a domain or module as herein described) or a portion that regulates expression of a product encoded by the polynucleotide (*e.g.*, a promoter). For example, the promoter that is naturally associated with the polynucleotide (*ie.* the *xabB* promoter) may be introduced into the cell that is the subject of testing. In this regard, where only the promoter is utilised, detecting modulation of the promoter activity can be achieved, for example, by operably linking the promoter to a suitable reporter polynucleotide including, but not restricted to, green fluorescent protein (GFP), luciferase, ß-galactosidase and catecholamine acetyl transferase (CAT). Modulation of expression may be determined by measuring the activity associated with the reporter polynucleotide.

In another example, the subject of detection could be a downstream regulatory or biosynthetic target of the target molecule, rather than target molecule itself or the reporter molecule operably linked to a promoter of a gene encoding a product the expression of which is regulated by the target protein.

These methods provide a mechanism for performing high throughput screening of putative modulatory agents such as proteinaceous or non-proteinaceous agents comprising synthetic, combinatorial, chemical and natural libraries. These methods will also facilitate the detection of agents which bind either the polynucleotide encoding the target molecule or which modulate the expression of an upstream molecule, which subsequently modulates the expression of the polynucleotide encoding the target molecule. Accordingly, these methods provide a mechanism of detecting agents that either directly or indirectly modulate the expression and/or activity of a gene or expression product according to the invention.

8. Production of secondary metabolites

The present invention further relates to a process for enhancing the level and/or functional activity of secondary metabolites, preferably albicidins, using one or more agents selected from the polynucleotides, polypeptides, fragments, variants, derivatives, vectors and modulatory agents described above. The process in a preferred embodiment, includes the steps of stably transforming a host cell with an expression vector as broadly described above, comprising at least one nucleic acid sequence encoding a polypeptide of the invention or a biologically active fragment or variant or derivative of these and isolating transformants which produce an enhanced amount of antibiotics, which are preferably of the albicidin class. The vector optionally comprises a signal sequence for secretion recognised by the host cell. Illustrative secretory leaders include the secretory leaders of penicillinase, α-factor, immunoglobulin, T-cell receptors, outer membrane proteins, glucoamylase, fungal amylase and the like. By fusion in proper reading frame, the mature polypeptide may be secreted into the medium. The host cell may be a eukaryote or a prokaryote cell. In one embodiment, the cell naturally produces polyketides, preferably antibiotic polyketides and, in this regard, the cell is preferably *X. albilineans* or other bacteria capable of producing albicidins. Optionally, the construct may include a transcription regulating sequence, which is not subject to repression by substances present

in the growth medium. The above process may be used to prepare antibiotics directly or they may be used to prepare cell free extracts containing increased quantities of antibiotics, preferably of the albicidin class, for *in vitro* preparation of said antibiotics. Suitably, these cell free extracts may be prepared for example using the method disclosed by Dobrogosz, W.J. (1981) Enzymatic activity. *In* Manual of Methods for General Bacteriology (Gerhardt, P., ed) Washington, DC: American Society for Microbiology, pp. 365-392. In a preferred embodiment, a vector from which a phosphopantetheinyl transferase (PPTase) can be translated is also introduced into the host cell. Expression of PPTase polynucleotides has been shown to be important for the production of polyketides in heterologous expression systems. Preferably, the PPTase is selected from EntD and/or XabA as for example disclosed herein. If desired, a vector from which a methyltransferase, more preferably and *O*-methyltransferase, and even more preferablv an *S*-adenosylmethionine *O*-methyltransferase can be translated may also be introduced into the host cell. An exemplary methyltransferase for this purpose is XabC as described herein.

Alternatively, the expression hosts may be used as a source of increased quantities of antibiotics, which can be subsequently purified as for example disclosed by Birch *et al.* in U.S. Patent No. 4,525,354.

The invention also contemplates use of the polynucleotides, polypeptides, fragments, variant and derivatives of the invention in methods of combinatorial biosynthesis of novel antibiotics as for example disclosed by Khosla *et al.* in U.S. Patent No. 5,712,146, Peterson *et al.* in U.S. Patent No. 5,783,431 and Betlach *et al.* in U.S. Patent No. 6,251,636 or in methods of producing antibiotics in hosts that ordinarily do not produce them as for example disclosed by Barr *et al.* in U.S. Patent No. 6,033,883. As discussed in Section 2.4, the invention contemplates albicidin PKS-NRPS derivatives with altered activities in one or more respects for the production of polyketides other than the albicidin natural product(s) of the XabB. In this regard, expression vectors containing nucleotide sequences encoding a variety of such derivatives for the production of different polyketides are transformed into the appropriate host cells to construct a library. In one embodiment, a mixture of such vectors is transformed into selected host cells and the resulting cells plated into individual colonies and selected to identify successful transformants. A variety of strategies is available to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the naturally occurring host gene cluster so

that each colony in the library produces a different PKS and ultimately a different polyketide, as for example disclosed by Betlach *et al.* in U.S. Patent No. 6,251,636. The libraries thus produced can be considered at four levels: (1) a multiplicity of colonies each with a different PKS-NRPS encoding sequence; (2) the proteins produced from the coding sequences; (3) the polyketides produced from the proteins assembled into a functional PKS-NRPS; and (4) antibiotics or compounds with other desired activities derived from the polyketides. Colonies in the library can be induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of polyketides. Polyketides that are secreted into the media or have been otherwise isolated can be screened for binding to desired targets, such as receptors, signalling proteins, and the like. The supernatants *per se* can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening methods involve detecting the binding of each member of the library to receptor or other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art. Alternatively, individual polyketide members of the library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included. Antibiotic activity can be verified using typical screening assays such as those for albicidin set forth in Example 1.

The invention also extends to the use of the polynucleotides, polypeptides, fragments, variant and derivatives of the invention for the synthesis of antibiotics, preferably antibiotics of the albicidin class.

The polynucleotides of the invention encoding XabB, or a biologically-active fragment or variant thereof, together with a recombinant polynucleotide encoding a PPTase and/or an *O*-methyltransferase which participate or which are capable of participating in the albicidin biosynthetic pathway, provide the means to engineer high level co-expression of the albicidin synthetase, its activating PPTase and modifying methyltransferase to obtain higher yields of albicidins.

In order that the invention may be readily understood and put into practical effect, particular preferred embodiments will now be described by way of the following non-limiting examples.

EXAMPLES

EXAMPLE 1

Albicidin multifunctional synthase gene

Materials and Methods

Bacterial strains and plasmids

The properties of bacteria and plasmids used in this example are listed in Table 1.

Media, culture conditions and antibiotics

X. albilineans strains were routinely cultured on SP medium (Birch & Patil, 1985b) at 28° C. *Escherichia coli* DH5α and JM109 were used as hosts in cloning experiments and were grown on LB medium at 37° C (Sambrook *et al.*, 1989). Broth cultures were aerated by shaking at 200 r.p.m. on an orbital shaker. Modified YEB medium (Van Larebeke *et al.*, 1977) for patch mating consisted of 10 mg ml^{-1} peptone, 5 mg mL^{-1} yeast extract, 5 mg mL^{-1} NaCl, 5 mg mL^{-1} sucrose and 0.5 mg mL^{-1} $MgSO_4.7H_2O$. The following antibiotics were added to media as required: 50 μg kanamycin mL^{-1}; 15 μg tetracycline mLl^{-1}; 100 μg ampicillin mL^{-1}.

Routine genetic procedures

Bacterial genomic DNA and plasmid DNA isolation, gel electrophoresis, DNA restriction digests, ligation reactions and transformation were performed by routine procedures (Sambrook *et al.*, 1989). DNA fragments were excised from agarose gels and residual agarose was removed with the BRESAclean™ DNA purification kit (GeneWorks, Adelaide).

Construction of a X. albilineans partial genomic library

Genomic DNA from *X. albilineans Xal*3 was digested with *Eco*RI and size-fractionated. DNA fragments of 15 to 20 kb were ligated to dephosphorylated *Eco*RI-cleaved pBluescript SK II. The ligated DNA was electroporated into E. coli TOP10.

Transformants were selected on LB agar medium containing ampicillin, and stored in LB broth with 15% glycerol at -70°C.

PCR amplification

*Bam*HI-digested genomic DNA from *X. albilineans* LS157 was religated at low concentration (0.5 μg/mL) to generate circular DNA molecules as templates for inverse PCR. Three primers, one from the IS terminal region of Tn5 (IR2: 5'-CGGGATCCTCACATGGAAG TCAGATCCTG-3'), and two flanking the unique BamHI restriction site of Tn5 (BL: 5'-GGGGACCTTGCACAGATAGC-3', and BR: 5'-CATTCCTGTAGCGGATGGAGATC-3'), were used to amplify the sequences flanking the Tn5 insertion in the genome of LS157. The amplified fragments (1.4-kb and 6.0-kb) were cloned into pZErO-2, yielding pZIL and pZIR (Figure 1).

PCR was performed in a volume of 50 μl with 200 ng of genomic DNA (or 10 ng of plasmid DNA), 0.4 ng/μL of each of primer, 0.2 mM of each dNTP, 1.8 mM Mg^{2+}, and 1 unit of elongase enzyme mix (Life Technologies). A 10-min initial denaturation step at 94° C was followed by 35 thermal cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 1 min per 1 kb of expected amplification.

Construction of promoter probes and glucuronidase assay

Plasmid pRG960sd contains a promoterless *β*-glucuronidase gene (*uidA*) downstream of a multiple cloning site (Van den Edde *et al.*, 1992). Sequence upstream of xabB (nucleotide residues 1005 to 1210 or 521 to 1210) was amplified from pLXABB by PCR. Forward primer P1F1 (5'-ACGCGGATCCCAGCAGGGTGTCATACACG-3'), or P1F2 (5'-TCGCGGATCC GCGCGATTGAAGTAGTCC-3') contained a BamHI restriction site (underlined). Reverse primer P1R (5'-TCCCCCGGGCGGCCAGCGTGGTGCTACTAC-3') introduced a *Xma*I restriction site (underlined). PCR fragments were ligated into *Bam*HI/*Xma*I-cut pRG960sd, yielding pRG960p1 and pRG960p2. These constructs were mobilised from E. coli DH5α into X. albilineans LS155 as described below.

Promoter strength was quantified by fluorometric analysis of glucuronidase activity (Jefferson, 1987; Xiao *et al.*, 1992). The protein content in cell lysates was determined by the dye-binding method (Bradford, 1976) using a Bio-Rad protein assay kit.

<u>Bacterial conjugation</u>

DNA transfer between *E. coli* donor (JM109 pLAFR3 ± insert, or DH5α pRG960sd ± insert) and *X. albilineans* recipient (LS157 or LS155) was accomplished by triparental transconjugation with helper strain pRK2013. Mid-log-phase cultures of the recipient were spotted onto agar plates containing YEB medium with no antibictics (20 μL per spot). After the liquid was absorbed by the agar, 20 μL of mid-log-phase culture of the helper was added to each spot. The liquid was again allowed to absorb, and 20 μL of mid-log-phase culture of the donor was added to each spot. After incubation of the mating plates overnight at 28° C, transconjugants were selected on SP plates supplemented with ampicillin, and tetracycline or spectinomycin.

<u>Assay and quantification of albicidin production</u>

Albicidin was quantified by a microbial plate bioassay as described previously (Birch and Patil, 1985b), except that the 10 mL basal layer of LB agar and the 5mL overlayer of 50% LB with 1% agar were supplemented with tetracycline or spectinomycin, and E. coli DH5α pLAFR3 or pRG960sd was used as the indicator strain. This change avoided interference by tetracycline or spectinomycin, which were added to some cultures to ensure retention of pLAFR3 or pRG960sd derivatives in *X. albilineans*. Inhibition zone widths in the bioassay were converted to albicidin concentrations by interpolation on a dose-response plot produced under the same assay conditions. The plot fits the formula: Log [Alb] = 0.3 W – 0.92, where [Alb] is units of albicidin per 20 μL sample assayed, and W is the width in millimetres of the zone of growth inhibition surrounding each well.

Results

<u>Cloning and sequencing of *xabB* gene required for albicidin production</u>

Xanthomonas albilineans Tox⁻ mutant LS157 contains a single Tn5 insertion, in a 4.1 kb *Cla*I restriction fragment or a 16.5 kb *Eco*RI restriction fragment (Figure 1).

Selection for kanamycin resistance, following shotgun cloning of *Cla*I restriction fragments of LS157 DNA into pBluescript II SK, yielded clone pBC157. Sequences flanking the Tn5 insertion in LS157 DNA were amplified by inverse PCR, and cloned into pZErO-2, producing pZIL and pZIR. Plasmid pLXABB was screened from *a X. albilineans Xa13 Eco*RI genomic library with probes described in Figure 1B. Subclones pSEBL and pSEBR were derived from pLXABB (Figure 1C, Table 1).

The double-strand sequence of the16,511 bp *Eco*RI genomic fragment in pLXABB was obtained by a primer-walking approach, using subclones pBC157, pZIL, pZIR, pSEBL, and pSEBR. The Tn5 insertion in the genome of LS157 is accompanied by 9-bp perfect repeat sequence (GTCCTGAAG), commencing at 2490 bp in GenBank accession no. AF239749.

The only ORF longer than 900 bp within the 16.5-kb fragment is disrupted by the Tn5 insertion. This ORF (designated xabB) encodes a protein of 4081 aa (Mr 525,695). It commences at 1230 bp in GenBank accession no. AF239749 with a TTG codon, 6 bp downstream from a ribosome binding sequence (RBS) GAGG, which may impose post-transcriptional control on the rate of gene product formation (McCarthy and Gualerzi, 1990). There is an alternative start codon (ATG) a further 15 bp downstream. Of the codons in this ORF, 8.5% are rarely used in E. coli. The closest match (TTGAGC-14x-TATAAC) to the consensus -35 (TTGACA) and -10 (TATAAT) sequences for *E. coli* σ^{70} promoters occurs 117 bp upstream of the translation initiation codon (Figure 2).

Downstream by 35 bp from the TAG stop codon of xabB is a probable RBS (GAGG), separated by 6 bp from the ATG start codon of another ORF (designated *xabC*) in the same orientation as *xabB*. Overlapping the *xabB* promoter region is another probable promoter for a divergent transcript including a putative RBS (TGGAGG) and start codon for a gene designated xatA, separated by 233 bp from *xabB* (Figure 1, 2).

<u>Complementation of *xabB* gene in LS157</u>

Mobilisation of pLAFR3, pLXABB1 or pLXABB2 by bacterial conjugation into Tox$^-$ mutant LS157 occurred at a frequency of 1.5×10^{-2} transconjugants/recipient cells. Albicidin production was undetectable in Tox$^-$ mutant LS157 and LS157 (pLAFR3)

controls, but introduction of the *xabB* gene on pLXABB1 or pLXABB2 restored albicidin production to the level of the wild-type parental strain LS155 (Figure 4).

Functional analysis of *xabB* promoter region

GUS activity was undetectable in LS155 and LS155 (pRG960sd) controls. Plasmid pRG960p1 or pRG960p2, with 206 bp or 690 bp from the *xabB* promoter region upstream of GUS, both conferred GUS activity with no difference in expression level or pattern in *X. albilineans* LS155 (Figure 5).

Discussion

Albicidin was partially characterised as a low-molecular-weight compound that contains 38 carbon atoms with 3-4 aromatic rings (Birch and Patil, 1985a). The compound is not degraded by peptidases (Birch and Patil, 1985a), but it is cleaved by the AlbD esterase (Zhang and Birch, 1997). Based on the deduced functionality of the synthase describe herein, albicidin is likely to be a complex polyketide, condensed with amino acid(s), or nonproteinogenic amino, hydroxyl and carboxyl acid(s) by C-N, amide or ester bond formation.

The characterisation of XabB as a multi-modular hybrid enzyme provides new insights into the mechanism of albicidin biosynthesis and possible approaches to engineer the overproduction of albicidins. For example, the complementation experiments (Figure 4) indicate that increased copy number of *xabB* stimulates early production of albicidin, but other factors become limiting during idiophase. It may be possible to increase expression of the albicidin synthase by modifications to the promoter and TTG start codon, or to improve albicidin yields by supplying candidate substrates (such as shikimate-derived units). The unusual enzyme organisation also contributes to the emerging understanding of how microbes generate structural diversity of antibiotics, and can facilitate combinatorial engineering of antibiotics of mixed peptide/polyketide origin.

EXAMPLE 2

Albicidin Antibiotic and Phytotoxin Biosynthesis in Xanthomonas albilineans *Requires a Phosphopantetheinyl Transferase Gene*

Materials and Methods

Bacterial strains and plasmids

The properties of bacteria and plasmids used in this Example are listed in Table 3.

Media, culture conditions and antibiotics

X. albilineans strains were routinely cultured on SP medium (Birch & Patil, 1985b) at 28° C. *Escherichia coli* DH5α and JM109 were used as hosts in cloning experiments and were grown on LB medium at 37° C (Sambrook *et al.*, 1989). Broth cultures were aerated by shaking at 200 r.p.m. on an orbital shaker. Modified YEB medium (Van Larebeke *et al.*, 1977) for patch mating consisted of 10 mg ml^{-1} peptone, 5 mg mL^{-1} yeast extract, 5 mg mL^{-1} NaCl, 5 mg mL^{-1} sucrose and 0.5 mg mL^{-1} $MgSO_4.7H_2O$. The following antibiotics were added to media as required: 50 μg kanamycin mL^{-1}; 15 μg tetracycline mLl^{-1}; 100 μg ampicillin mL^{-1}.

Assay of albicidin production

Albicidin was quantified by a microbial plate bioassay as described previously (Birch and Patil, 1985b), except that the 10 mL basal layer of LB agar and the 5 mL overlayer of 50% LB with 1% agar were supplemented with tetracycline, and E. coli DH5α [pLAFR3] was used as the indicator strain. This change avoided interference by tetracycline, which was added to some cultures to ensure retention of pLAFR3 lerivatives in *X. albilineans*.

Routine genetic procedures

Bacterial genomic DNA and plasmid DNA isolation, gel electrophoresis, DNA restriction digests, ligation reactions and transformation were performed by routine procedures (Sambrook *et al.*, 1989). DNA fragments were excised from agarose gels and

residual agarose was removed with the BRESAclean™ DNA purification kit (GeneWorks, Adelaide).

DNA sequencing and analysis

Sequencing reactions were performed by dideoxynucleotide chain termination (Sanger *et al.*, 1977) using the BigDye™ Terminator Cycle Sequencing Kit and 373A DNA sequencer (PE Applied Biosystems) through the Australian Genome Research Facility. Oligonucleotide primers were purchased from GeneWorks (Adelaide). University of Wisconsin Genetics Computer Group (UWGCG) programs BLASTP, FASTA, PILEUP, and BESTFIT were used through WebANGIS version 2.0 for DNA and protein sequence analyses of the GenBank, EMBL, PIR and SWISSPROT databases using standard defaults.

Cloning of Tn5 flanking sequences

*Eco*RI-digested genomic DNA from *X. albilineans* Tox^- mutant LS156 was ligated into pBluescript II SK and electroporated into *E. coli* DH5α. Transformants were selected on LB medium containing kanamycin and ampicillin, yielding clone pBEA1, from which subclones pCEA1 and pPEA1 were obtained (Figure 1).

Amplification of sequences from wild-type LS155 by PCR

Sequences flanking the Tn5 insertion in LS156 were used to design primers (A1F: 5'-TTTGGGTTGGATCGGGTAG-3' and A1R: 5'-CCTTCTCGTCCTTG CTCTTC-3') for PCR-amplification of the corresponding wild type *X. albilineans* LS155 chromosomal DNA. PCR was performed in a volume of 50 μL with 200 ng of genomic DNA, 0.4 ng μL^{-1} of each of primer, 0.2 mM of each of dNTP, 1.8 mM Mg^{2+}, and 1 unit of elongase enzyme mix (Life Technologies). A 4-min initial denaturation step at 94° C was followed by 35 thermal cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 2 min. The amplified DNA fragment was cloned into pGEM-T to give pGTA1 (Figure 1).

Construction of expression vectors

The coding region of the *xabA* gene was amplified from pGTA1 by PCR. Primer A1F1 (5'-GGAATTCCATGCCCAATGCCGTACCG-3') contained an *Eco*RI restriction

site (underlined) for insertion of the amplified gene into the correct reading frame of *lacZ* in pLAFR3. Primer A1R1 (5'-CGGGATCCCGTGCTCACCAGGCGTAGTGG-3') introduced a *Bam*HI restriction site (underlined), 5 bases downstream from the stop codon of the amplified gene. The amplified DNA fragment was digested with *Eco*RI and *Bam*HI, and ligated with *Eco*RI/*Bam*HI-digested pLAFR3 to result in pLXABA.

Similarly, the coding region of the *entD* gene was PCR-amplified from *E. coli* DH5α by colony PCR using primers EntDF (5'-TCCCGGAATTCCATGGTCGATATGAAAACTACGC-3') and EntDR (5'-GCCCAAGCTTCTAATCGTGTTGGCACAGCGTTATG-3'), then ligated into pLAFR3 to produce pLENTD. The inserts in pLXABA and pLENTD were sequenced to confirm the expected clones.

Bacterial triparental mating

DNA transfer between *E. coli* donor (JM109 pLAFR3 ± insert) and *X. albilineans* recipient (LS155 or LS156) was accomplished by triparental transconjugation with helper strain pRK2013. The mid-log-phase cultures of the recipient were spotted onto agar plates containing YEB medium with no antibiotics (20 μL per spot). After the liquid was absorbed by the agar, 20 μL of mid-log-phase culture of the helper was added to each spot. The liquid was again allowed to absorb, and 20 μl of mid-log-phase culture of the donor was added to each spot. After incubation of the mating plates overnight at 28° C, transconjugants were selected on SP plates supplemented with tetracycline and ampicillin.

Results

Cloning and sequencing of the xabA gene required for albicidin production

Xanthomonas albilineans Tox⁻ mutant LS156 contains a single Tn5 insertion, in a 3.0-kb *Eco*RI restriction fragment (Wall & Birch, 1997). Selection for Tn5-encoded kanamycin resistance, following shotgun cloning of *Eco*RI restriction fragments of LS156 DNA into pBluescript II SK, yielded pBEA1 (Figure 8).

Both strands of the insert in pBEA1 excluding the Tn5 insertion were sequenced by primer-walking from T3 and T7 vector sequences in pBEA1 and subclones pCEA1 and

pPEA1. The corresponding genomic region was amplified from wild-type *X. albilineans* LS155 by PCR, and cloned into pGEM-T to give pGTA1. Sequencing of pGTA1 revealed that a 9-bp imperfect repeat sequence (TTGGCCACG) in the genome of LS156 accompanied the Tn5 insertion (following base number 1869 in Figure 9). The double-strand nucleotide sequence of the 2989 bp wild type *Eco*RI fragment is deposited in GenBank under accession no. AF191324.

Reading frame analysis of the 3 kb *Eco*RI fragment revealed that only one ORF (designated xabA) is disrupted by the Tn5 insertion. This ORF encodes a protein of 278 aa (Mr 29 277), with 6.12% codons rarely used in *E. coli*. There were no close matches to *E. coli* -10 (TATAAT) and -35 (TTGACA) consensus promoter sequences, and no appropriately spaced RBS sequence (such as AGGA or GAGG) in the region upstream of the putative start codon ATG (Figure 9). A region of GC-rich dyad symmetry with a free energy of –10.2 kcals/mol was found, followed by two TCTC boxes that closely resemble the TCTG consensus sequence characteristic of many factor-independent termination sites (Brendel & Trifonov, 1984; Platt, 1986) downstream of the TGA termination codon of xabA.

Comparison of XabA with other bacterial PPTases

A search for proteins with homology to the deduced *xabA* product, using the FASTA and BLASTP and SWISSPROT programs, indicated regions of similarity to EntD from *Escherichia coli* (170 aa overlap, 35.9 % identity, 56.5 % similarity), *Shigella flexneri* (180 aa overlap, 35.0 % identity, 55.6 % similarity), *Salmonella typhimurium* (184 aa overlap, 35.9 % identity, 62.0 % similarity), and *Salmonella austin* (172 aa overlap, 36.1 % identity, 61.1 % similarity). XabA contains (V/I)G(V/I)D and (F/W)(S/C/T)xKE(S/A)xxK domains characteristic of the phosphopantetheinyl transferase (PPTase) superfamily, and shares 17-36 % overall identity, 39-62 % overall similarity, with other bacterial PPTases (Table 4).

Enhanced expression of xabA by complementation in LS156 results in increased production of albicidins

Mobilisation of pLAFR3 or pLXABA (pLAFR3::*xabA*) by triparental matings into Tox$^-$ mutant LS156 occurred at a frequency of 1.5 x 10^{-2} transconjugants/recipient

cells. Albicidin production was undetectable in Tox⁻ mutant LS156 and LS156 (pLAFR3) controls, but introduction of the *xabA* gene on pLXABA enhanced albicidin production restored albicidin production (Figure 10). In LS156 (pLXABA), as in LS155, albicidin was first detectable in late-log-phase cultures (OD_{550} = 0.7) and was maximal in stationary phase. Albicidin production was not responsive to IPTG or glucose, and the *lac* promoter driving *xabA* in pLXABA is considered to express constitutively in *X. albilineans*. The *E. coli entD* gene, expressed from the *lac* promoter in pLENTD, also complemented the *xabA*::Tn5 mutation, restoring albicidin production in LS156.

Discussion

A gene required for albicidin production in *X. albilineans* was isolated using a Tn5 mutagenesis and shotgun cloning approach. The ORF interrupted by Tn5 in Tox^- mutant LS156 is designated *xabA*. This ORF was isolated from Tox^+ parent strain LS155, and shown to enhance albicidin production early in the production phase in LS156 when expressed from the *lac* promoter in pLAFR3. Tn5 insertions typically cause polar mutations affecting all downstream cistrons in an operon (De Bruijn and Lupski, 1984). Complementation of the mutation in LS156 by the isolated *xabA* ORF indicates the absence of any downstream cistron involved in albicidin production. There is no consensus RBS sequence close to the alternative start codons for this ORF in the *X. albilineans* genome. Translation may be initiated without an evident ribosome binding sequence complementary to the 3' end of the 16S rRNA, as observed for some streptomycete genes involved in secondary metabolism (Strohl, 1992), and for some chloroplast genes (Kozak, 1999).

PPTases play an essential role in priming polyketide, fatty acid, non-ribosomal peptide and siderophore biosynthesis (Gehring *et al.*, 1997a; Lambalot *et al.*, 1996; Marahiel *et al.*, 1997; Walsh *et al.*, 1997). All polyketide synthase, fatty acid synthetases, and non-ribosomal peptide synthetases require post-translational modification to become catalytically active (Walsh *et al.*, 1997). The inactive apo-proteins are converted to their active holo-forms by transfer of the 4'-phosphopantetheinyl (P-pant) moiety of coenzyme A to the sidechain hydroxyl of a serine residue in a conserved carrier domain (Lambalot *et al.*, 1996; Walsh *et al.*, 1997). The P-pant moiety serves to covalently tether the growing

product, which is assembled by sequential action of multiple catalytic domains in these complex synthetases (Walsh *et al.*, 1997).

A family of more than twenty PPTases is recognised by a common (V/I)G(V/I)Dx40-45...(F/W)(S/C/T)xKE(A/S)xxK signature sequence, but overall sequence homologies are low (Gehring *et al.*, 1997; Lambalot *et al.*, 1996; Nakano *et al.*, 1992; Quadri *et al.*, 1998a). In *E. coli*, there are two PPTases with distinct specificities: ACPS is active on acyl carrier protein (ACP) domains in fatty acid and polyketide synthase; EntD is active on peptidyl carrier protein (PCP) and aryl carrier protein (ArCP) domains in peptide synthetases (Lambalot *et al.*, 1996; Walsh *et al.*, 1997). Thus, PPTases may be partner-protein specific. However, Sfp from *B. subtilis* appears to be non-specific, efficiently activating both fatty acid, polyketide synthase and peptide synthetases (Kealey *et al.*, 1998; Mofid *et al.*, 1999; Quadri *et al.*, 1998a). XabA includes the PPTase VGID and FSxKESxxK motifs. Although it has highest overall similarity to the peptide-selective EntD proteins, the sequence groupings are not sufficiently compelling to predict the specificity of XabA for polyketide synthase or peptide synthetases (Table 4, Figure 11).

Complementation studies have revealed substantial functional interchangeability of PPTases in different bacteria. For example, the *B. sublitis sfp* gene involved in surfactin biosynthesis complements mutants in *E. coli entD* (enterobactin biosynthesis) and *B. brevis gsp* (gramicidin biosynthesis) (Borchert *et al.*, 1994; Grossman *et al.*, 1993). In vitro, ACPS from *E. coli* activates apoproteins from *Lactobacillus, Rhizobium* and *Streptomyces* (Lambalot *et al.*, 1996). Because XabA shows highest similarity to EntD, we amplified the *entD*-coding region from *E. coli*, and arranged it for expression from the *lac* promoter in broad host-range vector pLAFR3. This construct (pLENTD) restored albicidin production in *X. albilineans xabA*::Tn5 mutant LS156. EntD is a peptide-selective PPTase that converts inactive apo-EntF and apo-EntB to active holo-enzymes involved in biosynthesis of enterobactin in *E. coli* (Gehring *et al.*, 1997a). Functional complementation of the *xabA*::Tn5 mutation by *entD* indicates that XabA is a PPTase required for post-translational activation of synthetases involved in albicidin production in *X. albilineans*. The specificity of EntD for activation of peptide synthetases in *E. coli* indicates that albicidin biosynthesis probably involves an XabA-activated peptide synthetase.

Some PPTase genes involved in non-ribosomally synthesised peptide biogenesis are located near the genes encoding their targets (Quadri *et al.*, 1998b). For example, *B. brevis gsp, B. sublitis sfp,* and *E. coli entD* genes all lie within 4 kb of operons encoding the target peptide synthetases (Borchert *et al.*, 1994; Coderre & Earhart, 1989; Nakano *et al.*, 1992). However, *M. tuberculosis pptT* is not located near the *mbt* gene cluster encoding the target peptide synthetases involved in mycobactin biosynthesis (Quadri *et al.*, 1998b). No gene encoding a PPTase has been identified in any of the antibiotic and phytotoxin biosynthetic gene clusters characterised from *Streptomyces* spp. (Gehring *et al.*, 1997b) and *Pseudomonas* spp. (Bender *et al.*, 1999). No evident target gene was found within 1282 bp upstream or 870 bp downstream of *xabA*. Three cosmids spanning about 100 kb in two regions of the genome complemented 56 of 58 tested Tox$^-$ mutants of *X. albilineans*, but not LS156 (Rott *et al.*, 1996). These results indicate that *xabA* is not clustered with the genes encoding the antibiotic synthetases that it activates.

Expression of *xabA* (or an alternative PPTase such as *entD*) is essential for albicidin biosynthesis. The phosphopantetheinyl transferase gene described herein provides new insight into antibiotic biosynthesis in the *Pseudomonadaceae*, and new opportunities to understand and apply albicidins as potent inhibitors of prokaryote DNA replication. This gene, together with the *xabB* provide the means to engineer high level co-expression of the albicidin synthetase and its activating PPTase to obtain higher yields of albicidins, and ultimately to manipulate the elements of this biosynthetic machinery, by mutagenesis or otherwise, to produce desired structural variants of this novel antibiotic class. They may also indicate a new approach to disease resistance, by engineering plants to interfere with the biosynthesis of albicidin toxins, which are key pathogenesis factors for the systemic development of leaf scald disease.

EXAMPLE 3

A methyltransferase gene is involved in albicidin biosynthesis in Xanthomonas albilineans

Material and Methods

Bacterial strains and plasmids

The properties of bacteria and plasmids used in this example are listed in Table 5.

Media, culture conditions and antibiotics

X. albilineans strains were routinely cultured on sucrose peptone (SP) medium at 28° C (Birch and Patil, 1985b). Escherichia coli strains were used as hosts in cloning experiments and were grown on LB medium at 37° C (Sambrook *et al.*, 1989). Broth cultures were aerated by shaking at 200 rpm on an orbital shaker. Modified YEB medium (Van Larebere *et al.*, 1977) was used for patch mating. The following antibiotics were added to media as required: kanamycin, 50 μg/mL; tetracycline, 15 μg/mL; ampicillin, 100 μg/mL.

Assay of albicidin production

Albicidin was quantified by a microbial plate bioassay as described previously (Birch and Patil, 1985b), except that the 10 mL basal layer of LB agar and the 5 mL overlayer of 50% LB with 1% agar were supplemented with tetracycline, and *E. coli* DH5α [pLAFR3] was used as the indicator strain. This change avoided interference by tetracycline, which was added to some cultures to ensure retention of pLAFR3 derivatives in *X. albilineans*.

Routine genetic procedures

Bacterial genomic DNA and plasmid DNA isolation, gel electrophoresis, DNA restriction digests, ligation reactions and transformation were performed by routine procedures (Sambrook *et al.*, 1989). DNA fragments were excised from agarose gels and residual agarose was removed with the BRESAclean™ DNA purification kit (GeneWorks, Adelaide).

DNA sequencing and analysis

Sequencing reactions were performed by dideoxynucleotide chain termination (Sanger *et al.*, 1977) using the BigDye™ Terminator Cycle Sequencing Kit and 373A DNA sequencer (PE Applied Biosystems) through the Australian Genome Research Facility. Oligonucleotide primers were purchased from GeneWorks (Adelaide). University of Wisconsin Genetics Computer Group (UWGCG) programs BLASTP, FASTA, PILEUP, and BESTFIT were used through WebANGIS version 2.0 for DNA and protein sequence analyses of the GenBank, EMBL, PIR and SWISSPROT databases.

Recovery of the downstream sequence of truncated *xabC* by IPCR

Genomic DNA of *X. albilineans* LS155 was digested with *Nco*I. Following phenol/chloroform extraction and ethanol precipitation, the digested DNA was self-ligated at a concentration of 0.5 µg/mL. The ligated DNA was precipitated with ethanol and resuspended in sterile H_2O to a concentration of 20 ng/µL as template for IPCR. Sequence of the 16.5 kb *Eco*RI fragment including the 5' region of *xabC* was used to design primers (IF: 5'-AAGCGTCGACATAGCAGCAG-3' and IR: 5'-CGGCAACGCATTCGACCTCG-3') for IPCR-amplification of the sequence downstream of the *Eco*RI site of truncated *xabC* gene.

IPCR was performed in a volume of 50 µL with 50 ng of template DNA, 0.4 ng/µL of each of primer, 0.2 mM of each of dNTP, 1.8 mM Mg^{2+}, and 1 unit of elongase enzyme mix with proof-reading activity (Life Technologies). A 10 min initial denaturation step at 94° C was followed by 35 thermal cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 1 min per 1 kb of expected amplification product. The IPCR product was cloned into pZErO-2 to give pZIXC. Clones of construct pZIXC from three independent PCR reactions were sequenced to rule out the possibility of PCR-generated errors.

Insertional mutagenesis

An internal 625 bp *Cla*I-*Eco*RI fragment of *xabC* (Figure 13) was firstly cloned into *Cla*I/*Eco*RI-digested pBluescript II SK to provide a *Kpn*I restriction site, then subcloned into *Eco*RI/*Kpn*I-cleaved pJP5603 to yield pJP-BEC. The inserts in pBluescript II SK intermediates (pBEC) were sequenced to confirm the expected clones.

The suicide construct pJP-BEC was transferred from the mobilising strain E. coli S17-1 (λpir) into *X. albilineans* LS155. Exconjugant colonies were selected on SP agar containing kanamycin and ampicillin. Insertional disruption in xabC or thp was verified by PCR using primers flanking the expected integration site of pJP-BEC or pJP-BAS and extension at 72° C for 1 min as previously described (Zhang and Birch, 1997b). The effect on albicidin biosynthesis was determined using the microbial plate assay. Representative (Tox^-) insertional mutants in *xabC* (LS-JP1) and *thp* (LS-JP2) were retained for further analysis.

Construction of expression vectors

The coding region of the *xabC* gene was amplified from *X. albilineans* LS155 chromosomal DNA by PCR. Primer A3F (5'-CGGGATCCCATGGATTCAGCGTTACC-3') contained a BamHI restriction site (underlined) for insertion of the amplified gene into the correct reading frame of *lacZ* in pLAFR3. Primer A3R (5'-CCCAAGCTTTCATTAT GGGGCCCTCTTGC-3') introduced a *Hin*dIII restriction site (underlined). The amplified DNA was digested with *Bam*HI and *Hin*dIII, and ligated with *Bam*HI/*Hin*dIII-digested pLAFR3 to result in pLXABC. *X. albilineans* Tox$^-$ mutant LS157 contains a single Tn5 insertion, in a 4.1 kb *Cla*I restriction fragment or a 16.5 kb EcoRI restriction fragment (Figure 12). Selection for kanamycin resistance, following shotgun cloning of *Cla*I restriction fragments of LS157 DNA into pBluescript II SK, yielded clone pBC157. Sequences flanking the Tn5 insertion in LS157 DNA were amplified by inverse PCR, and cloned into pZErO-2, producing pZIL and pZIR. The double-strand sequence of the 16,511 bp *Eco*RI genomic fragment in pLXABB was obtained by a primer-walking approach, using subclones pBC157, pZIL, pZIR, pSEBL, and pSEBR. The Tn5 insertion in the genome of LS157 is accompanied by 9-bp perfect repeat sequence (GTCCTGAAG), commencing at 2490 bp in GenBank accession no. AF239749.

Genetic complementation of albicidin biosynthesis

DNA transfer between E. coli donor (JM109 pLAFR3 ± insert) and *X. albilineans* recipient (LS-JP1 or LS-JP2), was accomplished by triparental transconjugation with helper strain pRK2013. Mid-log-phase cultures of the recipient were spotted onto agar plates containing YEB medium with no antibiotics (20 μL per spot). After the liquid was absorbed by the agar, 20 μL of mid-log-phase culture of the helper was added to each spot. The liquid was again allowed to absorb, and 20 μL of mid-log-phase culture of the donor was added to each spot. After incubation of the mating plates overnight at 28° C, transconjugants were selected on SP plates supplemented with ampicillin, and tetracycline or spectinomycin.

Transconjugants were tested for albicidin production using the microbial plate bioassay. The constructs pLXABB, pLXABC were designed to test complementation in trans. However, complementation could also occur in *cis*, by homologous recombination between the complementing construct and the insertionally mutated chromosomal gene. To

exclude this possibility, the retention of the insertion in *xabC* was confirmed by PCR, using primers from *aphA* (in the insertion) and *xabB* (adjoining *xabC* in the chromosome).

Results and Discussion

Cloning and sequencing of the full-length *xabC* gene

Downstream by 45 bp from the TAG stop codon of *xabB* is the start of an ORF (designated *xabC*) in the same orientation. The 639-bp sequence downstream of the *Eco*RI site of the truncated *xabC* was amplified from wt *X. albilineans* LS155 using IPCR. The double-strand nucleotide sequence of 1515 bp from the stop codon of *xabB* to the *Nco*I site downstream of *xabC* (Figure 13) is deposited in GenBank under accession no. AF239750. The *xabC* ORF encodes a protein of 343 aa (Mr 37,704). One TCTG-like sequence (TGTG) and one typical TCTG box characteristic of many factor independent termination sites (Brendel and Trifonov, 1984) occur downstream of the termination codon (TAA) of xabC (Fig. 2). However, the other features typical of such terminators (a region of GC rich dyad symmetry, followed by a run of consecutive thymine residues) are not present within 435 bp downstream of the *xabC* stop codon.

XabC is similar to O-methyltransferases

The deduced product of xabC shows 22-30% overall identity and 52-60% overall similarity to a family of methyltransferases that utilise S-adenosyl-methionine (SAM) as a co-substrate for O-methylation of small molecules (Ingrosso *et al.*, 1989; Haydock *et al.*, 1991; Kagan and Clarke, 1994). These enzymes include tetracenomycin polyketide C-8 O-methyltransferase (TcmO, P39896) and C-3 O-methyltransferase (TcmN, P16559) of *Streptomyces glaucescens*, hydroxyneurosporene-O-methyltransferase (P17061) of *Rhodobacterium capsulatus*, and hydroxyindole-O-methyltransferases of rat pineal and retina (O09179) and chicken pineal gland (Q92056). Three highly conserved motifs in SAM-dependent methyltransferases are also present in XabC as shown in Figures 13 and 14. The crystal structure analysis for the methyltransferase-SAM complex (Schlukebier *et al.*, 1995) provides firm structural evidence for the role of motif I in SAM binding.

Insertional mutagenesis of *xabC* blocks albicidin biosynthesis

Insertional mutation in *xabC* was accomplished using suicide-vector pJP-BEC and confirmed by PCR. Six out of eight tested transconjugants were verified by PCR to contain insertional mutations in *xabC*. Albicidin production was undetectable in these insertional mutants, compared to wt *X. albilineans* LS155 control. The other transconjugants may result from integration of the vector at other genomic locations by illegitimate recombinations as reported previously (Penfold and Pemberton, 1992).

Complementation test

Introduction of the *xabC* gene in pLXABC or the truncated *xabC* gene in pLXABB into insertional mutant LS-JP2 restored albicidin production to the level of the wt parental strain LS155. This indicates that *xabC* is essential for albicidin production in *X. albilineans*. The truncated *xabC* in pLXABB (SEQ ID NO: 106) encodes 277 residues (SEQ ID NO: 107), including all of the three conserved motifs of SAM-methyltransferases, and appears fully functional by complementation. The continued presence of an insertion in the chromosomal locus was confirmed by PCR. Thus, complementation was operating in *trans*. This also indicates that no other cistron downstream of *xabC* is required for albicidin production, because insertional mutagenesis typically causes polar mutations affecting all downstream cistrons in an operon (De Bruijn and Lupski, 1989).

Enhanced expression of *xabC* results in increased production of albicidins

Derivatives of *X. albilineans* strain LS155, in which an *xabC* gene, or fragment thereof, was introduced in *trans*, were tested for production of albicidin using the bioassay described above. The results, presented in Figure 15, show that expression of *xabC* cloned into pLAFR3 in derivatives of *X. albilineans* strain LS155 complements an insertional mutation in the chromosomal *xabC*, and also enhances albicidin production early in the production phase. Expression of the first part of the *xabB* operon, including the full-length *xabB* and a truncated but functional *xabC*, further enhances albicidin production.

The disclosure of every patent, patent application, and publication cited herein is hereby incorporated herein by reference in its entirety.

The citation of any reference herein should not be construed as an admission that such reference is available as "Prior Art" to the instant application

Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. Those of skill in the art will therefore appreciate that, in light of the instant disclosure, various modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the present invention. All such modifications and changes are intended to be included within the scope of the appended claims.

TABLES

TABLE 1

Bacterial strains, and plasmids for Example 1

Strain or plasmids	*Relevant characteristics*	*Reference or source*
Strains		
E. coli		
DH5α	Φ80dlacZΔM15, Δ(lacZYA-argF	Promega
JM109	[F', lacI^{q}ZΔM15], Δ(lac-proAB	Promega
TOP10	F$^-$, Δ(mrr-hsdRMS-mcrBC), Δ(are-leu)7697, ΔlacX74	Invitrogen
X. albilineans		
Xa13	Wild-type albicidin producer from sugarcane (Queensland), Ap^r	Inventor's laboratory
LS155	Wild-type albicidin producer from sugarcane (Queensland), Ap^r	Wall and Birch (1997)
LS157	LS155::Tn5, albicidin deficient (Tox$^-$), Km^r St^r Ap^r	Wall and Birch (1997)
Plasmids		
pBluescript II SK	ColE1 origin, E. coli cloning vector, Ap^r	Stratagene
pZErO-2	ColE1 origin, E. coli cloning vector, Km^r	Invitrogen
pRK2013	ColE1origin, IncP, Tra$^+$, helper plasmid, Km^r	Ditta et al (1980)
pLAFR3	RK2 origin, Tra$^-$, Mob$^+$, broad host-range cosmid, Tc^r	Stachelhaus at al. (1987)
pRG960sd	ColE1origin, broad host-range plasmid, contains promoterless uidA with start codon and Shine-Dalgarno sequence, Sm^r Sp^r	Van den Edde *et al.* (1992)
pBC157	9.9-kb ClaI fragment carrying Tn5 and flanking sequences from LS157, in pBluescript II SK, Km^r Ap^r	This study
pZIL	1.4-kb fragment, inverse PCR amplified from LS157 in pZErO-2, Km^r	This study
pZIR	6.0-kb fragment, inverse PCR amplified from LS157 in pZErO-2, Km^r	This study
pZTI	0.9-kb fragment, PCR amplified from LS157 in pZErO-2, Km^r	This study
pXABB	16.5-kb EcoRI fragment from Xa13 in pBluescript II SK, Ap^r	This study
pSEBL	7.9-kb EcoRI-SpeI frament from pXABB in pBluescript II SK, Ap^r	This study
pSEBR	8.6-kb EcoRI-SpeI frament from pXABB in pBluescript II SK, Ap^r	This study

Strain or plasmids	*Relevant characteristics*	*Reference or source*
pLXABB1	16.5-kb EcoRI fragment from pXABB in pLAFR3 (xabB in the same direction as lac), Tc^r	This study
pLXABB2	16.5-kb EcoRI fragment from pXABB in pLAFR3 (xabB in the opposite direction to lac), Tc^r	This study
pRG960p1	206-bp BamHI-XmaI frament in pRG960sd, Sm^r Sp^r	This study
pRG960p2	690-bp BamHI-XmaI frament in pRG960sd, Sm^r Sp^r	This study

TABLE 2

Comparison of conserved sequences in peptide synthetases and XabB

Domain	*Core*	*Sequence conserved in peptide synthetases*[a]	*Sequence in XabB*	*Position in Xab (aa)*
Adenylation	A1	L(T/S)YxEL	WSYAQL	3806-3811
	A2	LKAGxAYL(V/L)P(L/I)D	FKAGACYVPID	3851-3861
	A3	LAYxxYTSG(S/T)TGxPKG	LACVMVTSGSTGRPKG	3917-3932
	A4	FDxS	FAVS	3967-3970
	A5	NxYGPTE	NNYGCTE	4063-4069
	A6	GELxIxGxG(V/L)ARGYL	GELHVHSVGMARGYW	4114-4128
	A7	Y(R/K)TGDL	YKTGDM	4152-4157
	A8	GRxDxQVKIRGxRIELGEIE	GRQDFEVKVRGHRVDTRQVE	4170-4189
	A9	LPxYM(I/V)P	LPTYMLP	4239-4245
	A10	NGK(V/L)DR	NGKLDR	4259-4264
Peptidyl carrier protein	PCP	DxFFxLGG(H/D)S(L/I)	DNFFALGGHSL MDFFAVGGHSV	4306-4316 3261-3271
Condensation	C1	SxAQxR(L/M) (W/Y)xL	TYAQERLWLV SLFQERLWFV	3333-3342 4374-4383
	C2	RHExLRTxF	RHEVLRTRF RHEILRTRF	3381-3389 4421-4429
	C3	MHHxISDG(W/V)S	IHHIISDGWS MHHLIYDAWS.	3456-3465 4498-4507
	C4	YxD(F/Y)AVW	YADYALW YADYAIW	3495-3501 4538-4544
	C5	(I/V)GxFVNT(Q/L) (C/A)xR	IGFFINILPLR IGFFINILPLR	3606-3617 4649-4659
	C6	(H/N)QD(Y/V)PFE	HQSVPFE NQALPFE	3641-3647 4685-4691
	C7	RDxSRNPL	RDSSQIPL RDTSRIPL	3658-3665 4701-4708

[a]*Sourced from reference (Marahiel et al., 1997).*

TABLE 3

Bacterial strains, and plasmids for Example 2

Strain or plasmids	*Relevant characteristics*	*Reference or source*
Strains		
E. coli		
DH5α	Φ80dlacZΔM15, recA1, endA1, gyrA96, thi-1, hsdR17(r_k^-, m_k^+) supE44, relA1, deoR, Δ(lacZYA-argF)U169	Promega
JM109	[F', traD36, proAB, lacI^{q}ZΔM15], recA1, endA1, gyrA96, thi hsdR17(r_k^-, m_k^+), supE44, relA1, Δ(lac-proAB)	Promega
X. albilineans		
Xa13	Wild-type albicidin producer from sugarcane (Queensland), Ap^r	This laboratory
LS155	Wild-type albicidin producer from sugarcane (Queensland), Ap^r	Wall & Birch (1997)
LS156	LS155::Tn5, albicidin deficient (Tox^-), Km^r St^r Ap^r	Wall & Birch (1997)
Plasmids		
pBluescript II SK	ColE1 origin, E. coli cloning vector, Ap^r	Stratagene
pGEM-T	ColE1 origin, E. coli TA-cloning vector, Ap^r	Promega
pRK2013	ColE1origin, IncP, Tra^+, helper plasmid, Km^r	Ditta et al. (1980)
pLAFR3	RK2 origin, Tra^-, Mob^+, broad host-range cosmid, Tc^r	Staskawicz et al. (1987)
pBEA1	8.8-kb EcoRI fragment carrying Tn5 and flanking sequences from LS156, in pBluescript II SK, Km^r Ap^r	This study
pCEA1	1766-bp *Eco*RI-*Cla*I fragment from pBEA1 in pBluescript II SK, Ap^r	This study
pPEA1	697-bp *Eco*RI-*Pst*I fragment from pBEA1 in pBluescript II SK, Ap^r	This study
pGTA1	2.1-kb fragment, PCR amplified from LS155 in pGEM-T, Ap^r	This study
pLXABA	834-bp *Eco*RI-*Bam*HI fragment (xabA ORF) from pGTA1 in pLAFR3, Tc^r	This study
pLENTD	630-bp *Eco*RI-*Hin*dIII fragment (entD ORF) from DH5□ in pLAFR3, Tc^r	This study

TABLE 4

Similarity between XabA and other PPTases involved in antibiotic and fatty acid biosynthesis in bacteria

Pathway	Protein	Organism	Specificity (A/P)†	Domain I Domain II	Homology (ID/SIM)
Albicidin	XabA	X.albilineans	?	GVGIDLERP--(X)39--FSAKESLFKAJY	-
Enterobactin	EntD	E.coli	P‡	PIGIDIEEI--(X)36--FSAKESAFKASE	35.9/56.5
		S.flexneri	?	PIGVDIEEI--(X)36--FSAKESAFKASE	35.0/55.6
		S.typhimurium	?	RIGIDIEKI--(X)35--FSAKESVYKAFQ	35.9/62.0
		S.austin	?	RVGVDIEKI--(X)35--FSAKESVYKAFQ	36.1/61.1
Mycobactin	PptT	M.tuberculosis	P	SVGIDAEPH--(X)34--FCAKEATYKAWF	30.5/55.5
Surfactin	Sfp	B.subtilis	A/P‡	PIGIDIEKT--(X)35--WSMKESFIKQEG	24.8/48.5
	Psf-1	B.pumilus	?	PVGIDIEEI--(X)35--WSMKEAFIKLTG	19.8/47.6
Gramicidin	Gsp	B.brevis	P‡	PVGIDIERI--(X)35--WTIKESYIKAIG	20.8/42.0
Iturin A	Lpa-14	B.subtilis	?	PIGIDIEKM--(X)35--WSMKESFIKQAG	20.0/43.4
Fatty acids	HI0152	H. influenzae	?	AVGIDIEFP--(X)34--WCLREAVLKSQG	19.7/45.7
	AcpS	E. coli	A‡	GLGTDIVEI--(X)40--FAVKEAAAKAFG	16.5/38.8
		M.tuberculosis	A	GVGIDLVSI--(X)41--WAAKEAVIKAWS	25.7/47.6
		B. subtilis	?	GIGLDITEL--(X)41--FAAKEAFSKAFG	25.5/46.2
PPTase domain*				(V/I)G(I/V)D (F/W)(S/C/T)XKE(S/A)XXK	

TABLE 5

Bacterial strains, and plasmids for Example 3

Strain or plasmids	*Characteristics*	*Reference or source*
Strain		
E. coli		
DH5α	Φ80dlacZΔM15, Δ(lacZYA-argF)U169	Promega
JM109	[F', lacI^{q}ZΔM15], Δ(lac-proAB)	Promega
TOP10	F$^-$, Δ(mrr-hsdRMS-mcrBC), Δ(are-leu)7697, ΔlacX74	Invitrogen
S17-1λpir	S17-1 lysogenized with λpir	Penfold and Pemberton (1992)
X. albilineans		
Xa13	wt albicidin producer from sugarcane (Queensland), Ap^r	Our laboratory
LS155	wt albicidin producer from sugarcane (Queensland), Ap^r	Wall and Birch (1997)
LS157	xabB::Tn5, albicidin deficient (Tox$^-$), Km^r St^r Ap^r	Wall and Birch (1997)
LS-JP1	thp::pJP-BAS, albicidin deficient (Tox$^-$), Km^{r}Ap^r	This work
LS-JP2	xabC::pJP-BEC, albicidin deficient (Tox$^-$), Km^{r}Ap^r	This work
Plasmids		
pBluescript II SK	ColE1 origin, E. coli cloning vector, Ap^r	Stratagene
pZErO-2	ColE1 origin, E. coli cloning vector, Km^r	Invitrogene
pRK2013	ColE1origin, IncP, Tra$^+$, helper plasmid, Km^r	Ditta *et al.* (1980)
pLAFR3	RK2 origin, Tra$^-$, Mob$^+$, broad host-range cosmid, Tc^r	Staskawicz *et al.* (1987)
pJP5603	Bacterial suicide vector, Km^r	Penfold and Pemberton (1991)
pZIXC	1 kb IPCR product in pZErO-2, Km^r	This work
pBAS	278 bp ApaI-SalI fragment of thp in pBluescrpt II SK, Ap^r	This work

Strain or plasmids	*Characteristics*	*Reference or source*
pJP-BAS	284 bp SalI-KpnI fragment from pBAS in pJP5606, Km^r	This work
pBEC	625 bp ClaI-EcoRI fragment of xabC in pBluescript II SK, Ap^r	This work
pJP-BEC	655 bp EcoRI-KpnI fragment from pBEC in pJP5603, Km^r	This work
pLTHP	1226 bp EcoRI-BamHI fragment from pLXABB in pLAFR3, Tc^r	This work
pLXABC	1029 bp xabC ORF amplified from LS155 in pLAFR3, Tc^r	This work
pLXABB	16.5 kb EcoRI fragment from Xa13 in pLAFR3, Tc^r	This work

BIBLIOGRAPHY

Albertini, A.M., Caramori, T., Scoffone, F., Scotti, C. and Galizzi, A. (1995) Sequence around the 159° region of the *Bacillus subtilis* genome: the *pksX* locus spans 33.6 kb. *Microbiol* **141:** 299-309.

Aparicio, J.F., Molnar, I., Schwecke, T., Konig, A., Haydock, S.F., Khaw, L.E., *et al.* (1996) Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169:** 9-16.

Bender, C. L., Alarcon-Chaidez, F. & Gross, D. C. (1999). *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* **63:** 266-292.

Birch, R.G., and Patil, S.S. (1983) The relation of blocked chloroplast differentiation to sugarcane leaf scald disease. *Phytopathology* **73:** 1368-1374.

Birch, R. G. & Patil, S. S. (1985a). *Antibiotic and process for the production thereof.* U.S. patent 4525354, 25 June 1985.

Birch, R.G., and Patil, S.S. (1985b) Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J Gen Microbiol* **131**: 1069-1075.

Birch, R.G., and Patil, S.S. (1987a) Correlation between albicidin production and chlorosis induction by *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Physiol Mol Plant Pathol* **30**: 199-206.

Birch, R.G., and Patil, S.S. (1987b) Evidence that an albicidin-like phytotoxin induces chlorosis in sugarcane leaf scald disease by blocking plastid DNA replication. *Physiol Mol Plant Pathol* **30**: 207-214.

Borchert, S., Stachelhaus, T. & Marahiel, M. A. (1994). Induction of surfactin production in *Bacillus subtilis* by gsp, a gene located upstream of the gramicidin S operon in *Bacillus brevis*. *J Bacteriol* **8:** 2458-2462.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:** 248-254.

Brendel, V. & Trifonov, E. N. (1984). A computer algorithm for testing potential prokaryotic terminator. *Nucleic Acids Res* **12:** 4411-4427.

Cane, D.E., Walsh, C.T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**: 63-68.

Coderre, P. E. & Earhart, C. F. (1989). The *entD* gene of the *Escherichia coli* K12 enterobactin gene cluster. *J. Gen Microbiol* **135:** 3043-3055.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**: 4174-4183.

Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., and Leadlay P.F. (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **348**(6297): 176-178.

De Bruijn, F. J. & Lupski, J. R. (1984). The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - a review. *Gene* **27:** 131-149.

Ditta G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980) Broad host range DNA cloning system for Gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc Natl Acad Sci USA* **77:** 7347-7351.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* **252:** 675-679.

Donadio S., and Katz L. (1992) Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin fonnation in Saccharopolyspora erythraea. *Gene* **111**(1): 51-60.

Donadio, S., Staver, M.J., 1993. IS136, an insertion element in the erythromycin gene cluster of *Saccharopolyspora erythraea*. *Gene* **126:** 147-151.

Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., *et al.* (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc Natl Acad Sci USA* **96:** 13294-13299.

Gehring, A. M., Bradley, K. A. & Walsh, C. T. (1997a). Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36:** 8495-8503.

Gehring, A. M., Lambalot, R. H., Vogel, K. W., Drueckhammer, D. G. & Walsh, C. T. (1997b). Ability of *Streptomyces* spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for *E. coli* holo-ACP synthase. *Chem Biol* **4:**17-24.

Gehring, A.M., DeMoll, E., Fetherston, J.D., Mori, I., Mayhew, G.F., Blattner, F.R., *et al.* (1998) Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem Biol* **5:** 573-586.

Gross, D.C., 1991. Molecular and genetic analysis of toxin production by pathovars of *Pseudomonas syringae*. *Ann. Rev. Phytopath.* **29:** 247-278.

Grossman, T. H., Tuckman, M., Ellestad, S. & Osburne, M. S. (1993). Isolation and characterization of *Bacillus subtilis* genes involved in siderophore biosynthesis: relationship between *B. subtilis sfp^o* and *Escherichia coli entD* genes. *J Bacteriol* **19:** 6203-6211.

Guenzi, E., Galli, G., Grgurina, I., Gross, D.C., and Grandi, G. (1998) Characterization of the syringomycin synthetase gene cluster: a link between prokaryotic and eukaryotic peptide synthetases. *J Biol Chem* **273:** 32857-32863.

Haydock, S.F., Dowson, J.A., Dhillon, N., Robert, G.A., Cortes, J., Leadlay, P.F., 1991. Cloning and sequencing analysis of genes involved in erythromycin biosynthesis in Saccharopolyspora erythraea: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol. Gen. Genet.* **230:** 120-128.

Hopwood, D.A. and Khosla, C., Secondary Metabolites: Their Function and Evolution (1992) Wiley Chichester (Ciba Foundation Symposium 171) pp. 88-112

Hopwood, D.A., and Sherman, D.H. (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* **24:** 37-66.

Hopwood, D.A. (1997) Genetic contributions to understanding polyketide synthases. *Chem Rev* **97:** 2465-2497.

Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) Crystal structure of β-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J* **17:** 1183-1191.

Ingrosso, D., Fowler, A. V., Bleibaum, J., Clarke, S., 1989. Sequence of the D-aspartyl L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs from protein, DNA, RNA and small molecule S-adenosylmethionine dependent methyltransferase. *J. Biol. Chem.* **264:** 20131-20139.

Jefferson, R.A. (1987) Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol Biol Rep* **5:** 387-405.

Kagan, R.M., Clarke, S., 1994. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferase suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* **310:** 417-427.

Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C. & Barr, P. J. (1998). Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc Natl Acad Sci USA* **95:** 505-509.

Keating, T.A., and Walsh, C.T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Curr Opin Chem Biol* **3:** 598-606.

Konz, D., and Marahiel, M.A. (1999) How do peptide synthetases generate structural diversity? *Chem Biol* **6:** R39-R48.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234:** 187-208.

Lambalot, R. H. & Walsh. C. T. (1995). Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J Biol Chem* **270:** 24658-24661.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., *et al.* (1996) A new enzyme superfamily – the phosphopantetheinyl transferases. *Chem Biol* **3:** 923-936.

MacNeil, D.J., Occi, J.L., Gewain, K.M., MacNeil, T., Gibbons, R.H., Ruby, C.L., and Danis, A.J. (1992) Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. *Gene* **115:** 119-125.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* **97**: 2651-2673.

Martin, J.F., Liras, P., 1989. Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Ann. Rev. Microbiol* **43:** 173-206.

McCarthy, J.E.G., and Gualerzi, C. (1990) Translational control of prokaryotic gene expression. *Trends Genet* **6:** 78-85.

Mofid, M. R., Marahiel, M. A., Ficner, R. & Reuter, K. (1999). Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases. *Acta Crys* D **55:** 1098-1100.

Motamedi, H., and Shafiee, A. (1998) The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur J Biochem* **256**: 528-534.

Nakano, M. M., Corbell, N., Besson, J. & Zuber, P. (1992). Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol Gen Genet* **232:** 313-321.

Paitan, Y., Alon, G., Orr, E., Ron, E.Z., and Rosenberg, E. (1999) The first gene in the biosynthesis of the polyketide antibiotic TA of Myxococcus xanthus codes for a unique PKS module coupled to a peptide synthetase. *J Mol Biol* **286**: 465-474.

Penfold, R.J., Pemberton, J.M., 1992. An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene* **118:** 145-146.

Platt, T. (1986). Transcription termination and the regulation of gene expression. *Ann Rev Biochem* **55:** 339-72.

Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. (1998a). Characterization of Sfp, a *Bacillus subtillis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37:** 1585-1595.

Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. (1998b). Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* **5:** 631-645.

Rangaswamy, V., Mitchell, R., Ullrich, M., Bender, C., 1998. Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. *J. Bacteriol.* **180:** 3330-3338.

Reimman, C., Moore, R., Little, S., Savioz, A., Willetts, N.S., Haas, D., 1989. Genetic structure, function and regulation of the transposable element IS21. *Mol. Gen. Genet.* **215:** 416-424.

Ricaud, C., and Ryan, C.C. (1989) Leaf scald, In *Disease of Sugar Cane: major disease.* Ricaud, C., Egan, B.T., Gillaspie Jr., A.G., and Hughes, C.G. (eds). Amsterdam: Elsevier Science Publisher B. V., pp. 39-53.

Rott, P.C., Costet, L., Davis, M.J., Frutos, R., and Gabriel, D.W. (1996) At least two separate gene clusters are involved in albicidin production by *Xanthomonas albilineans. J Bacteriol* **178**: 4590-4596.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Sanger, F., Nickles, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463-5467.

Schluckebier, G., O'Gara, M., Sanger, W., Cheng, X.D., 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247:** 16-20.

Schmid, S., Seitz, T., Haas, D., 1998. Cointegrase, a naturally occurring truncated form of IS21 transposase, catalyzes replicon fusion rather than simple insertion of IS21. *J. Mol. Biol.* **282:** 571-583.

Scrutton, N.S., Berry, A., and Perham, R.N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343:** 38-43.

Solinas, F., Marconi, A.M., Ruzzi, M., Zennaro, E., 1995. Characterization and sequence of a novel insertion sequence, IS1162, from Pseudomonas flurescens. *Gene* **155:** 77-82.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis - catalytic role of the condensation domain. *J Biol Chem* **273:** 22773-22781.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6:** 493-505.

Staskawicz, B., Dahlbeck, D., Keen, N., and Napoli, C. (1987) Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Psendomonas syringae* pv. *glycinea. J Bacteriol* **169:** 5789-5794.

Strohl, W. R. (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res* **20:** 961-974.

Tang, L., Yoon, Y.J., Choi, C., and Hutchinson, C.R. (1998) Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei. Gene* **216:** 255-265.

Turgay, K., Krause, M., and Marahiel, M.A. (1992) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6:** 529-546.

Van den Edde, G., Deblaere, R., Goethals, K., Van Montagu, M., and Hosters, M. (1992) Broad host range and promoter selection vectors for bacteria that interact with plants. *Mol Plant-Microbe Interact* **5:** 228-234.

Van Larebeke, N., Genetello, C., Hernalsteens, J.P., Depicker, A., Zaenen, I., Messens, E., *et al.* (1977) Transfer of Ti plasmids between *Agrobacterium* strains by mobilization with the conjugative plasmid RP4. *Mol Gen Genet* **152:** 119-124.

Wall, M.K., and Birch, R.G. (1997) Genes for albicidin biosynthesis and resistance span at least 69 kb in the genome of *Xanthomonas albilineans. Lett Appl Microbiol* **24:** 256-260.

Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E. & Flugel, R. S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr Opin Chem Biol* **1:** 309-315.

Xiao, Y., Lu, Y., Heu, S., and Hutcheson, S.W. (1992) Organization and environmental regulation of the *Pseudomonas syringae* pv. *syringae* 61 *hrp* cluster. *J Bacteriol* **174:** 1734-1741.

Zhang, L., and Birch, R.G. (1997) The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. *Proc Natl Acad Sci USA* **94:** 9984-9989.

Zhang, L., Birch, R. G., 1997b. Mechanism of biocontrol by *Pantoea dispersa* of sugar cane leaf scald disease caused by *Xanthomonas albilineans. J. Appl. Microbiol.* **82:** 448-454.

Zhang, L., Xu, J., and Birch, R.G. (1998) Factors affecting biosynthesis by *Xanthomonas albilineans* of albicidin antibiotics and phytotoxins. *J Appl Microbiol* **85:** 1023-1028.

Zhang, L., Xu, J., and Birch, R.G. (1999) Engineered detoxification confers resistance against a pathogenic bacterium. *Nature Biotech* **17:** 1021-1024.

CLAIMS

1. An isolated polypeptide comprising at least one domain selected from the group consisting of:

(a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.

(b) a β-ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;

(c) a β-ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;

(d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;

(e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

(f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, and variants thereof; and

(g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

2. The polypeptide of claim 1, wherein the AL domain comprises each of the sequences set forth in SEQ ID NO: 6 and 8, or variants thereof.

3. The polypeptide of claim 1, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 10, 12 and 14, or variants thereof.

4. The polypeptide of claim 1, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 16, 18 and 20, or variants thereof.

5. The polypeptide of claim 1, wherein the A domain comprises each of the sequences set forth in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

6. The polypeptide of claim 1, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 54, 56, 58, 60, 62, 64 and 66, or variants thereof.

7. The polypeptide of claim 1, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

8. The polypeptide of claim 1, wherein the domains are arranged in an N- to C-terminal direction as follows: AL-ACP-KS-KR-ACP-ACP-KS-PCP-C-A-PCP-C.

9. The polypeptide of claim 1, comprising the sequence set forth in SEQ ID NO: 2, or biologically active fragment thereof, or variant or derivative of these.

10. The polypeptide of claim 9, wherein the variant has at least 60% sequence identity to the sequence set forth in SEQ ID NO: 2.

11. The polypeptide of claim 9, wherein the biologically active fragment is at least 6 amino acids in length.

12. An isolated polypeptide comprising at least a biologically active fragment of the sequence set forth in SEQ ID NO: 2 or variant or derivative thereof.

13. The polypeptide of claim 12, wherein the biologically active fragment is at least 6 amino acids in length.

14. The polypeptide of claim 12, wherein the biologically active fragment comprises at least one domain selected from the group consisting of:

(a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.

(b) a β-ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;

(c) a β-ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;

(d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;

(e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

(f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, and variants thereof; and

(g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

15. The polypeptide of claim 13, wherein the AL domain comprises each of the sequences set forth in SEQ ID NO: 6 and 8, or variants thereof.

16. The polypeptide of claim 13, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 10, 12 and 14, or variants thereof.

17. The polypeptide of claim 13, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 16, 18 and 20, or variants thereof.

18. The polypeptide of claim 13, wherein the A domain comprises each of the sequences set forth in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

19. The polypeptide of claim 13, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 54, 56, 58, 60, 62, 64 and 66, or variants thereof.

20. The polypeptide of claim 13, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

21. The polypeptide of claim 12, wherein the variant has at least 60% sequence identity to said at least a biologically active fragment.

22. The polypeptide of claim 12, wherein the variant has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 or 80.

23. An isolated polypeptide comprising at least biologically active fragment of the sequence set forth in SEQ ID NO: 83, or a variant or derivative thereof.

24. The polypeptide of claim 23, wherein the biologically active fragment comprises at least one of the consensus PPTase sequence motifs set forth in SEQ ID NO: 89 or 93, or variant thereof.

25. The polypeptide of claim 24, wherein the biologically active fragment comprises both the consensus PPTase sequence motifs set forth in SEQ ID NO: 89 or 93, or variant thereof.

26. The polypeptide of claim 23, wherein the biologically active fragment comprises the intervening sequence between said consensus PPTase sequence motifs, which intervening sequence comprises the sequence set forth in SEQ ID NO: 91, or variant thereof.

27. The polypeptide of claim 23, wherein the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof.

28. The polypeptide of claim 23, wherein the biologically active fragment is at least 6 amino acids in length.

29. The polypeptide of claim 23, wherein the variant has at least 60% sequence identity to the sequence set forth in SEQ ID NO: 83.

30. The polypeptide of claim 23, wherein the variant has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 87, 89, 91 or 93.

31. An isolated polypeptide comprising at least biologically active fragment of the sequence set forth in SEQ ID NO: 95, or a variant or derivative thereof.

32. The polypeptide of claim 31, wherein the biologically active fragment comprises at least one of the consensus methyltransferase sequence motifs set forth in SEQ ID NO: 99, 101 or 103, or variant thereof.

33. The polypeptide of claim 31, wherein the biologically active fragment comprises all the consensus methyltransferase sequence motifs set forth in SEQ ID NO: 99, 101 and 103, or variant thereof.

34. The polypeptide of claim 31, wherein the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 105, or variant thereof.

35. The polypeptide of claim 31, wherein the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 107, or variant thereof.

36. The polypeptide of claim 31, wherein the biologically active fragment is at least 6 amino acids in length.

37. The polypeptide of claim 31, wherein the variant has at least 60% sequence identity to the sequence set forth in SEQ ID NO: 95.

38. The polypeptide of claim 31, wherein the variant has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 99, 101 or 103.

39. An isolated polynucleotide comprising a sequence encoding at least one domain selected from the group consisting of:

(a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.

(b) a β-ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;

(c) a β-ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;

(d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;

(e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

(f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, and variants thereof; and

(g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

40. The polynucleotide of claim 39, wherein the AL domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 5 or 7, or variants thereof.

41. The polynucleotide of claim 40, wherein the AL domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 5 and 7, or variants thereof.

42. The polynucleotide of claim 39, wherein the KS domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 9, 11, 13, 15, 17 and 19, or variants thereof.

43. The polynucleotide of claim 42, wherein the KS domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 9, 11 and 13, or variants thereof.

44. The polynucleotide of claim 42, wherein the KS domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 15, 17 and 19, or variants thereof.

45. The polynucleotide of claim 39, wherein the KR domain is encoded by a nucleotide sequence set forth in SEQ ID NO: 21, or variant thereof.

46. The polynucleotide of claim 39, wherein the ACP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 23, 25 and 27, or variants thereof.

47. The polynucleotide of claim 39, wherein the A domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof.

48. The polynucleotide of claim 47, wherein the A domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof.

49. The polynucleotide of claim 39, wherein the PCP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 49 and 51, or variants thereof.

50. The polynucleotide of claim 39, wherein the C domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79, or variants thereof.

51. The polynucleotide of claim 50, wherein the C domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 53, 55, 57, 59, 61, 63 and 65, or variants thereof.

52. The polynucleotide of claim 50, wherein the C domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 67, 69, 71, 73, 75, 77 and 79, or variants thereof.

53. The polynucleotide of claim 39, comprising the sequence set forth in any one of SEQ ID NO: 1 or 3, or a biologically active fragment thereof at least 18 nucleotides in length, or a polynucleotide variant of these.

54. The polynucleotide of claim 53, wherein the polynucleotide variant has at least 60% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 1 or 3.

55. The polynucleotide of claim 53, wherein the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 1 or 3 under at least low stringency conditions.

56. The polynucleotide of claim 39, wherein the polynucleotide variant comprises a nucleotide sequence encoding at least one said domain.

57. The polynucleotide of claim 56, wherein the nucleotide sequence variant has at least 60% sequence identity to any one or more of the sequences set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79.

58. The polynucleotide of claim 56, wherein the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79 under at least low stringency conditions.

59. An isolated polynucleotide comprising a sequence encoding at least biologically active fragment of the sequence set forth in SEQ ID NO: 83, or a variant or derivative thereof.

60. The polynucleotide of claim 59, comprising the sequence set forth in any one of SEQ ID NO: 82 and 84, or a biologically active fragment thereof, or a polynucleotide variant of these.

61. The polynucleotide of claim 59, comprising a contiguous sequence of nucleotides at least 18 nucleotides in length and contained within the sequence set forth in SEQ ID NO: 86, or variant thereof.

62. The polynucleotide of claim 59, wherein the polynucleotide variant has at least 60% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 82, 84 and 86.

63. The polynucleotide of claim 59, wherein the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 82, 84 and 86 under at least low stringency conditions.

64. The polynucleotide of claim 59, wherein the polynucleotide variant comprises a nucleotide sequence encoding at least one PPTase sequence motif selected from SEQ ID NO: 89 and 93, or variant thereof.

65. The polynucleotide of claim 64, wherein the polynucleotide variant comprises a nucleotide sequence encoding the intervening sequence between the said consensus PPTase sequence motifs, said nucleotide sequence comprising the sequence set forth in SEQ ID NO: 91.

66. The polynucleotide of claim 59, wherein the polynucleotide variant suitably comprises a nucleotide sequence encoding a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof.

67. The polynucleotide of claim 66, wherein the contiguous sequence is encoded by the sequence set forth in SEQ ID NO: 86, or nucleotide sequence variant thereof displaying at 60% identity thereto.

68. The polynucleotide of claim 64, wherein the PPTase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 88 and 92, or nucleotide sequence variant thereof displaying at 60% identity thereto.

69. The polynucleotide of claim 65, wherein the intervening sequence is encoded by the nucleotide sequence set forth in SEQ ID NO: 90, or nucleotide sequence variant thereof displaying at 60% identity thereto.

70. The polynucleotide of claim 66, wherein the contiguous sequence is encoded by the sequence set forth in SEQ ID NO: 86, or nucleotide sequence variant thereof displaying at 60% capable of hybridising thereto under at least low stringency conditions.

71. The polynucleotide of claim 64, wherein the PPTase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 88 and 92, or nucleotide sequence variant thereof capable of hybridising thereto under at least low stringency conditions.

72. The polynucleotide of claim 65, wherein the intervening sequence is encoded by the nucleotide sequence set forth in SEQ ID NO: 90, or nucleotide sequence variant thereof capable of hybridising thereto under at least low stringency conditions.

73. An isolated polynucleotide comprising a sequence encoding at least biologically active fragment of the sequence set forth in SEQ ID NO: 95, or a variant or derivative thereof.

74. The polynucleotide of claim 73, comprising the sequence set forth in any one of SEQ ID NO: 94 and 96, or a biologically active fragment thereof, or a polynucleotide variant of these.

75. The polynucleotide of claim 73, comprising a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 104, or variant thereof.

76. The polynucleotide of claim 73, comprising a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 106, or variant thereof.

77. The polynucleotide of claim 73, wherein the polynucleotide variant has at least 60% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 94, 96, 104 and 106.

78. The polynucleotide of claim 73, wherein the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 94, 96, 104 and 106 under at least low stringency conditions.

79. The polynucleotide of claim 73, wherein the polynucleotide variant comprises a nucleotide sequence encoding a methyltransferase sequence motif selected from any one or more of SEQ ID NO: 99, 101 and 103, or variant thereof.

80. The polynucleotide of claim 79, wherein the methyltransferase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 98, 100 and 102, or nucleotide sequence variant thereof displaying at least 60% identity thereto.

81. The polynucleotide of claim 79, wherein the methyltransferase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 98, 100 and 102, or nucleotide sequence variant thereof capable of hybridising thereto under at least low stringency conditions.

82. An expression vector comprising the polynucleotide of any one of claims 39, 59 or 73, wherein the polynucleotide is operably linked to a regulatory polynucleotide.

83. A host cell containing the expression vector of claim 82.

84. A multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression vector for the production of the polypeptide of claim 1 or claim 12.

85. A method for enhancing the level and/or functional activity of an albicidin, said method comprising:

– introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of the polypeptide of claim 1 or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of the polypeptide of claim 1 or variant or derivative thereof can be translated;

– and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin.

86. The method of claim 85, further comprising introducing into said host cell a vector from which a PPTase can be translated.

87. The method of claim 86, wherein the PPTase is selected from EntD or XabA.

88. The method of claim 85, further comprising introducing into said host cell a vector from which a methyltransferase can be translated.

89. The method of claim 86, wherein the methyltransferase is XabC.

90. An antigen-binding molecule that is immuno-interactive with the polypeptide of claim 1 or claim 12.

91. An antigen-binding molecule that is immuno-interactive with the polypeptide of claim 23.

92. An antigen-binding molecule that is immuno-interactive with the polypeptide of claim 31.

93. A method of preparing a polynucleotide encoding a modified PKS, comprising using a nucleotide sequence encoding the polypeptide of claim 1 or claim 12 as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement.

94. A method for producing polyketides, comprising expressing the modified albicidin PKS encoding nucleotide sequence produced by the method of claim 93 in a suitable host cell to thereby produce a polyketide different from that produced by said polypeptide.

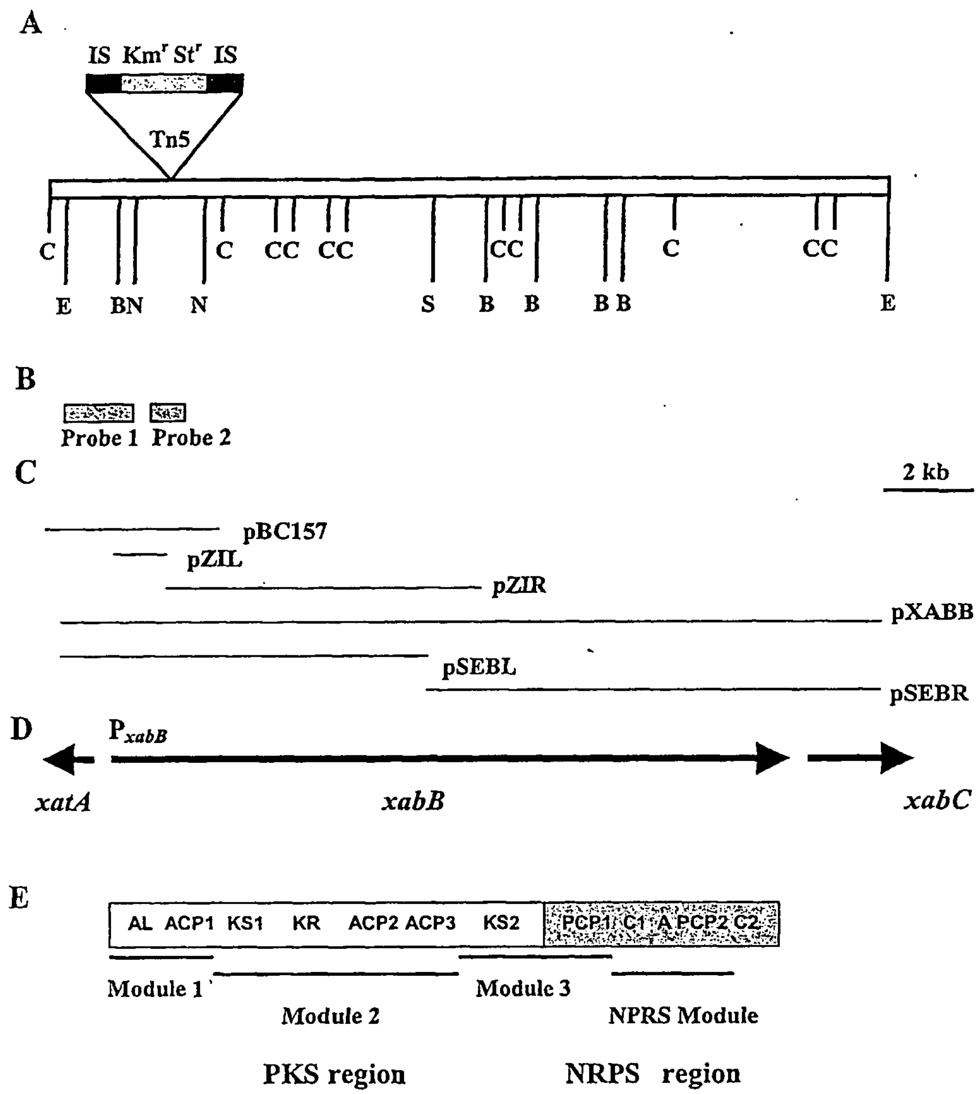

A
IS Km^r St^r IS
Tn5
C
C
CC
CC
CC
C
CC
E
BN
N
S
B
B
B B
E
B
Probe 1 Probe 2
C
2 kb
pBC157
pZIL
pZIR
pXABB
pSEBL
pSEBR
D
P$_{xabB}$
xatA
xabB
xabC
E
AL ACP1 KS1 KR ACP2 ACP3 KS2 PCP1 C1 A PCP2 C2
Module 1
Module 2
Module 3
NPRS Module
PKS region
NRPS region

FIGURE 1

```
                                                      P1F1
967   GGCGATGCGC GCACACTGCA GGTCCATCAC GCCACCTCCA GCAGGGTGTC
      CCGCTACGCG CGTGTGACGT CCAGGTAGTG CGGTGGAGGT CGTCCCACAG
      A  I  R  A  C  Q  L  D  M  M          RBS
                 xatA <------------

1017  ATACACGGCC AGCGGATGCT GCAGGTTTTC CACTGGCAGG GCCACTGGCT
      TATGTGCCGG TCGCCTACGA CGTCCAAAAG GTGACCGTCC CGGTGACCGA

                            -35 (P_xabB)         -10 (P_xabB)
1067  GTCGTAAGGG AAGCGGTGCC TTGAGCGCCG GTGCGGACAG TATAACGACA
      CAGCATTCCC TTCGCCACGG AACTCGCGGC CACGCCTGTC ATATTGCTGT
                                                -10 (P_xatA)

1117  CGTTCCTTGG CCAAGCGCAC TGTCGGCACG GCCTTGCTGA TGCCGCCCAT
      GCAAGGAACC GGTTCGCGTG ACAGCCGTGC CGGAACGACT ACGGCGGGTA
          -35 (P_xatA)

1167  GTAGCCGCGC GCCTGGATCT CGCGTAGTAG CACCACGCTG GCCGGGATCC
      CATCGGCGCG CGGACCTAGA GCGCATCATC GTGGTGCGAC CGGCCCTAGG
                                         P1R

       RBS           ------> xabB
1217  ATCGAGGGCG CGCTTGCCCA ATGCGCTCAT GCAGATAACT CTTGTAGCCG
      TAGCTCCCGC GCGAACGGGT TACGCGAGTA CGTCTATTGA GAACTACGGC
                    M  P  N  A  L  M  Q  I  T  L  V  A
```

FIGURE 2

(i). AL

```
TSGSSGESKGILLSH--GYFRTGDL  Xal-XabB(AL)
TGGTTGVAKGAMLTH--GWMATGDI  Hin-LCFA
TSGSTGTPKAVMLNH--GWFETGDL  Bsu-PksJ
SSGSTGDPKGVMLTH--GWVKTGDL  Bsu-MycA(AL)
SSGTTGLPKGVMLTH--GWLHTGDI  Pcr-ComL2
TSGTTGRPKGVVSAQ--GWYRTGDL  Sma-FkbB(AL)
TSGTTGRPKGVVSTQ--GWFRTGDL  Ame-RifA(AL)
TSGTTGTPKGVLSTQ--GWYRTGDL  Shy-RapA(AL)
```

(ii). KS

```
GPSEVINSACSSSLVAL--VELHGTGTSL--ALGHLGAAAG  Xal-XabB (KS1)
GPSLAVDTACSASLTAI--IEAHGTGTVL--NIGHAESAAG  Xal-XabB (KS2)
GPSLFVHTNCSSSLSAL--VEAHGTGTLL--NLGHLDTVAG  Mxa-Ta1
GPAVTVDTACSSSLVAV--IEAHGTGTKL--NIGHLFEAAG  Bsu-MycA
GPAVTVDTACSSSLVAL--VEAHGTGTRL--NIGHAQAAAG  Ser-EryA1
GPAMTVDTACSSGLTAL--VEAHGTGTRL--NIGHTQAAAG  Ser-EryA3
GPSVLVDTACSGGLTAL--VECHGTGTQA--NIGHLEGASG  Che-PKS1
GPSLAVDTACSSSLTAI--LEAHGTGTAL--NIGHCESAAG  Bsu-PksM
GPSVAVDTACSSSLVAI--VEAHGTGTLL--NLGHTEAAAG  Mtu-PpsA
GPSLTIDTACSSSLMAL--VEAHGTGTKV--NMGHPEPASG  Chick-FAS
GPSIALDTACSSSLLAL--IEAHGTGTKV--NMGHPEPASG  Rat-FAS
          *              *             *
```

(Active site cysteine) (Active site histidine)

(iii). KR

```
VYVVIGGAGGLGEVLSEHLIRTYD.AQLIWIGR  Xal-XabB
VYVISGGTGALARLFVAEIGKRATRATVILVAR  Mxa-Ta1
TVLVTGGTGGVGGQIARWLARRG.APHLLLVSR  Ser-EryA1
TVLVTGGTGGIGAHLARWLARSG.AEHLVLLGR  Ser-EryA3
SYLLVGGVGGLGSATALAMSTRG.ARHLLLINR  Che-PKS1
SYIITGGLGGLGLFFASKLAAAG.CGRIVLTAR  Mtu-MAS
SYIITGGLGGFGLELAQWLIERG.AQKLVLTSR  Chick-FAS
SYIITGGLGGFGLELARWLVLRG.AQRLVLTSR  Rat-FAS
```

(iv). ACP

```
CELALDSLQCVR  Xal-XabB(ACP1)
EYYGVDSIVAIE  Xal-XabB (ACP2)
ESYGVDSIVIIE  Xal-XabB (ACP3)
IGFGLDSIMLTQ  Bsu-MycA
ERYGIDSIIITQ  Mxa-Ta1
AELGVDSLSALE  Ser-EryA1
QDYGIDSLVAVE  Che-PKS1
IEYGLDSLGMLE  Mtu-MAS
ADLGLDSLMGVE  Chick-FAS
ADLGLDSLMGVE  Rat-FAS
      *
```

* (Active site serine)

FIGURE 3

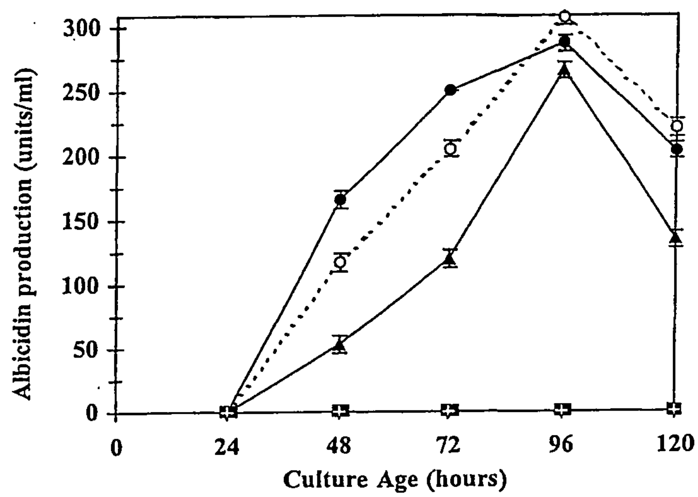

Albicidin production (units/ml)
300
250
200
150
100
50
0
0
24
48
72
96
120
Culture Age (hours)

FIGURE 4

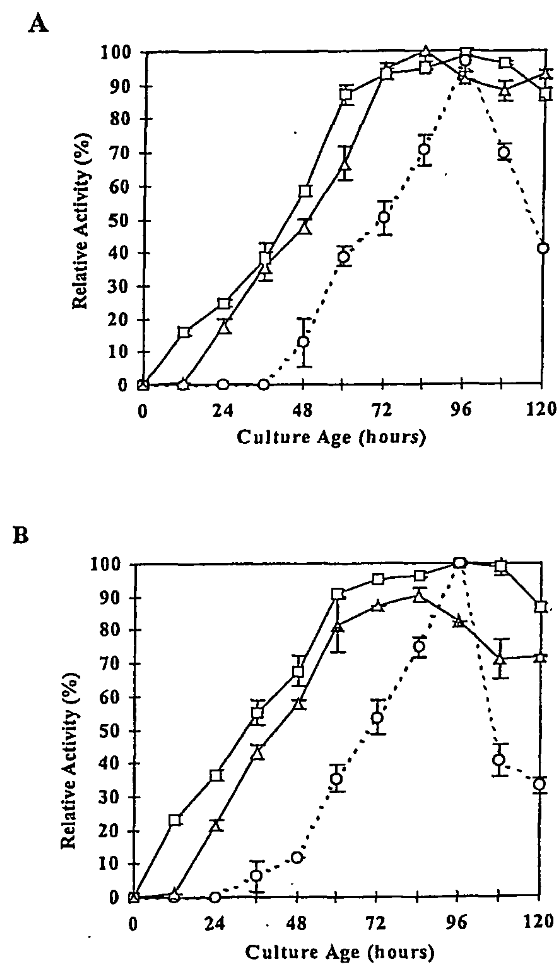
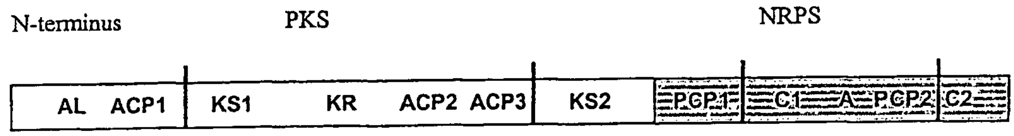

A
100
90
80
70
60
50
40
30
20
10
0
Relative Activity (%)
0
24
48
72
96
120
Culture Age (hours)
B
100
90
80
70
60
50
40
30
20
10
0
Relative Activity (%)
0
24
48
72
96
120
Culture Age (hours)

FIGURE 5

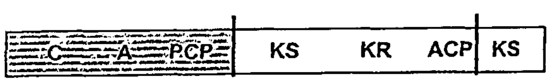
A. *X. albilineans* XabB (4801 aa)



B. *B. subtilis* MycA (3971 aa)

C. *Yersinia pestis* HMWP1 (3163 aa)

D. *M. xanthus* Ta1 (2392 aa)



E. *B. subtilis* PksorfX6 (4447 aa)

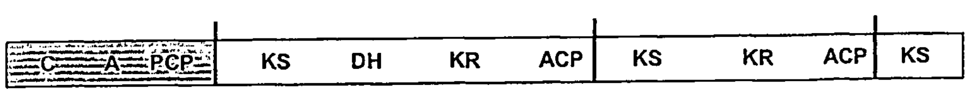


FIGURE 6

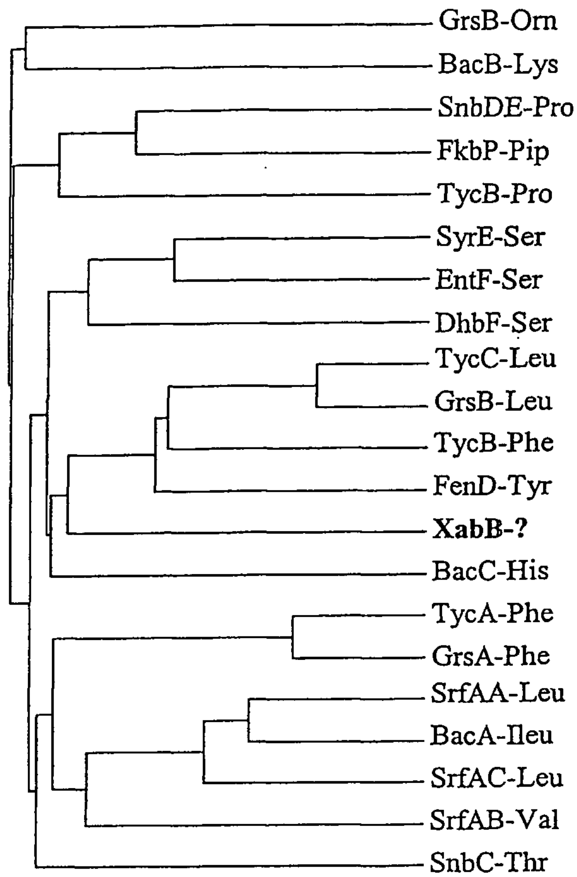

GrsB-Orn
BacB-Lys
SnbDE-Pro
FkbP-Pip
TycB-Pro
SyrE-Ser
EntF-Ser
DhbF-Ser
TycC-Leu
GrsB-Leu
TycB-Phe
FenD-Tyr
XabB-?
BacC-His
TycA-Phe
GrsA-Phe
SrfAA-Leu
BacA-Ileu
SrfAC-Leu
SrfAB-Val
SnbC-Thr

FIGURE 7

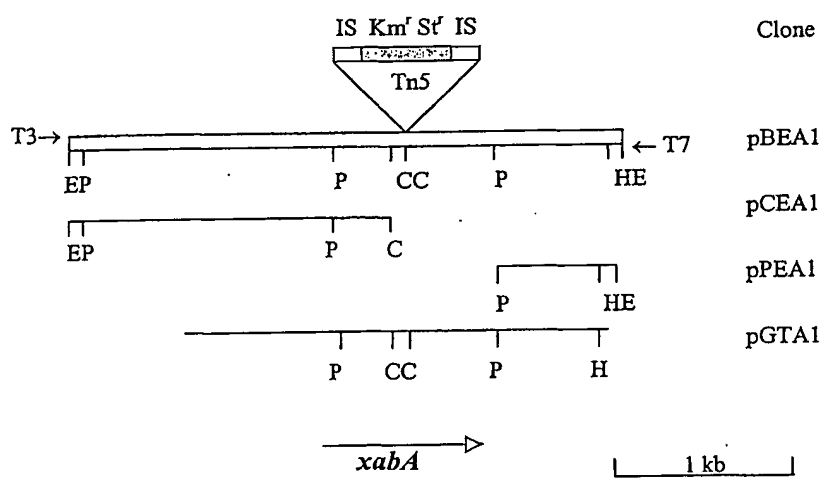

IS Km^r St^r IS
Tn5
Clone
T3→
← T7
pBEA1
EP
P
CC
P
HE
pCEA1
EP
P
C
pPEA1
P
HE
pGTA1
P
CC
P
H
xabA
1 kb

FIGURE 8

```
1054 TTCCCGCCCGAATAGGCGCAGGAAGCCAATAAGTATGGCAGCGCCCTTGACCAATGACAAGCTCATGCACCCAGGACGCCC

1135 GCTCTGCTCCGCGTCGTCCATCGCCATTGCGCCCCTCCCCGACCCCAAGCATCGACCAAAGGACCGAATGCGGCGGGTAGG

1216 CGCGACTCTGCGACACTAGCGCAATGTTATCGTCGACATTGACGCCCACAGCCCTCAGCGCAACGCAATGCCCAATGCCGT
                                -10                        RBS
                                                                       M  P  N  A  V  5
1297 ACCGATGCAGGGCGCGCGGGGACTCCCGCAGCCGCAAGCGATGAACCCAGGGTTGCCGAGCGTCGGCGGCTTGAGCGCAGG
      P  M  Q  G  A  R  G  L  P  Q  P  Q  A  M  N  P  G  L  P  S  V  G  G  L  S  A  G  32

1378 CCAGCCATTGCAGTTGTCGTTAGCACCGGAACTGCAGGCAGCCGCGCGCAGTGCCCACCGCCATCTGCTCGACGACGGCAC
      Q  P  L  Q  L  S  L  A  P  E  L  Q  A  A  A  R  S  A  H  R  H  L  L  D  D  G  T  59

1459 GGCGCTTTACCTGCTGGCGTTCGATACCGCGCAATTCGACCCGGGGGCTTTCGCGGCAATGGCAATCGCCCGCCCGGACAG
      A  L  Y  L  L  A  F  D  T  A  Q  F  D  P  G  A  F  A  A  M  A  I  A  R  P  D  S  86

1540 CATCGCCCGCAGCGTGCGCAAGCGTCAGGCCGAGTTCCTGTTCGGCCGTCTGGCCGCGCGACTGGCGCTGCAAGAGGTGCT
      I  A  R  S  V  R  K  R  Q  A  E  F  L  F  G  R  L  A  A  R  L  A  L  Q  E  V  L  113

1621 GGGACCTGCGCAAGCGCAGGCAGATATTGCAATCGGCGCGACGCGCGCGCCCTGCTGGCCTGCCGGCAGCCTGGGCAGCAT
      G  P  A  Q  A  Q  A  D  I  A  I  G  A  T  R  A  P  C  W  P  A  G  S  L  G  S  I  140

1702 TTCCCATTGCGAGGACTACGCGGCCGCCATCGCCATGGCGGCCGGCACCCGCCACGGCGTGGGCATCGATCTGGAACGACC
      S  H  C  E  D  Y  A  A  A  I  A  M  A  A  G  T  R  H  G [V  G  I  D] L  E  R  P  167

1783 AATCACACCCGCGGCGCGCGCGGCGTTGCTGAGCATCGCAATCGATGCCGACGAAGCCGCTCGTCTGGCAAAGGCGGCAGA
      I  T  P  A  A  R  A  A  L  L  S  I  A  I  D  A  D  E  A  A  R  L  A  K  A  A  D  194

          ∇Tn5
1864 CGCGCAGTGGCCGCAAGACCTGCTGCTGACCGCACTATTTTCGGCCAAGGAAAGCCTGTTCAAAGCCGCCTACAGCGCGGT
      A  Q  W  P  Q  D  L  L  L  T  A  L [F  S] A [K  E  S] L  F [K] A  A  Y  S  A  V  221

1945 CGGACGCTACTTCGACTTCAGCGCGGCACGCCTGTGCGGCATCGACCTGGCACGGCAATGCCTGCATCTGCGCCTGACCGA
      G  R  Y  F  D  F  S  A  A  R  L  C  G  I  D  L  A  R  Q  C  L  H  L  R  L  T  E  248

2026 GACACTCTGCGCGCAATTCGTGGCCGGGCAAGTGTGCGAGGTCGGCTTCGCGCGCCTACCACCGGACCTGGTGCTCACCCA
      T  L  C  A  Q  F  V  A  G  Q  V  C  E  V  G  F  A  R  L  P  P  D  L  V  L  T  H  275

2107 CTACGCCTGGTGAGCACGCGGACAGTCGAACCCGCCAACGCCAACGGCACTCAAGACGTGGCGTGCGCCGCGTCGGTCGTG
      Y  A  W  *                                                                        278

2188 AAGCTCTCCCCGCAGCCGCACTCGGCGGTGGCATTGGGATTGCGGAACACGAAGGTCTCACCCAAGCCCTGCTTGGCGAAG

2269 TCGATTTCGGTGCCATCGACCAACTGCAGACTGGCGGCATCGACATAAATCCGCACTCCGTCCTGCTCGAACACCGCATCG

2350 TCCGCGCGTGCCTCGTGCGCCAGATCGGTGACATGGCCCCAACCGGAACAGCCTGTGCGTACCACCCCGAAACGTAGACCC
```

FIGURE 9

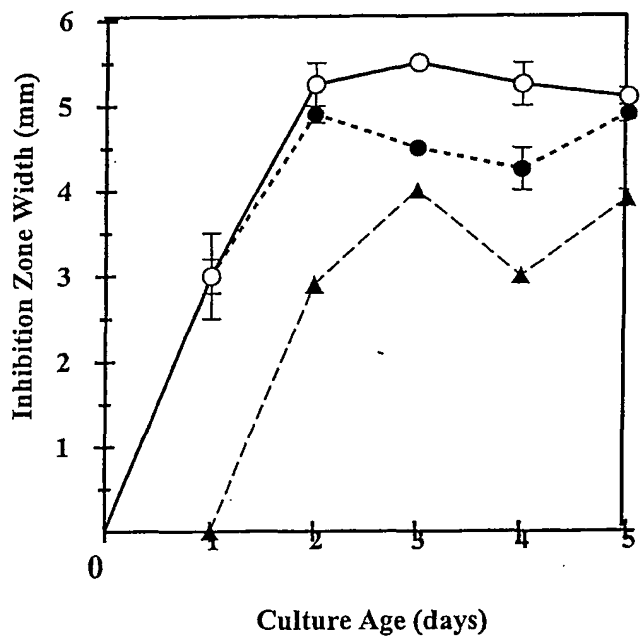


FIGURE 10

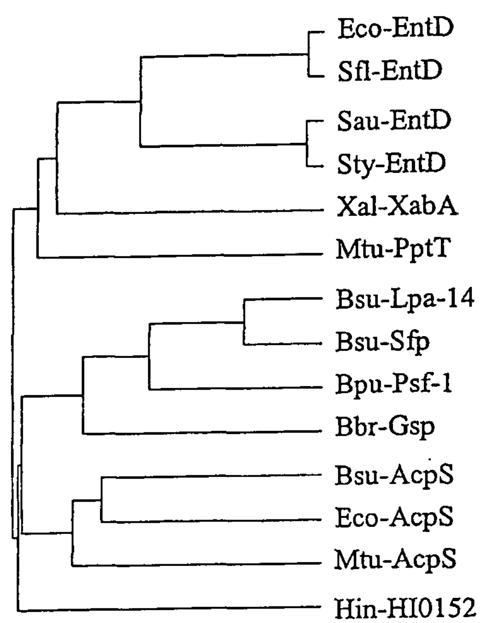

Eco-EntD
Sfl-EntD
Sau-EntD
Sty-EntD
Xal-XabA
Mtu-PptT
Bsu-Lpa-14
Bsu-Sfp
Bpu-Psf-1
Bbr-Gsp
Bsu-AcpS
Eco-AcpS
Mtu-AcpS
Hin-HI0152

FIGURE 11

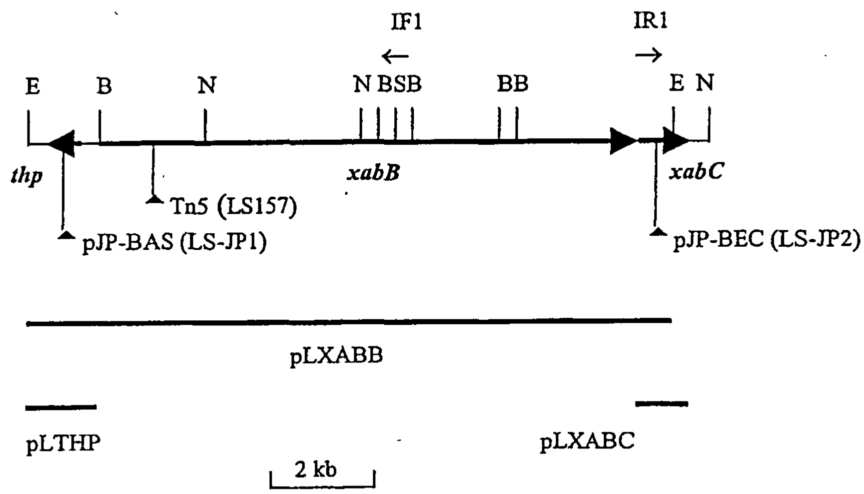

IF1
IR1
E
B
N
N BSB
BB
E N
thp
xabB
xabC
Tn5 (LS157)
pJP-BAS (LS-JP1)
pJP-BEC (LS-JP2)
pLXABB
pLTHP
pLXABC
2 kb

FIGURE 12

+1 ClaI RBS (A3F)

HincII

xabB stop codon
TAGCAAAAAGCCGGCCGCCGTCACCCGTTCATCGATAGCGAGGGCAATCATGGATTCAGCGTTACCTACATCTGCATTTACCTTCGATCTCTTTTACACCACGGTTAAC 105
xabC M D S A L P T S A F T F D L F Y T T
V N 20

GCCTACTATCGCACTGCCGCAGTCAAGGCGGCGATCGAACTGGGGCTATTCGATGTGGTGGGGCAGCAGGGCCGAACTCCCGCAGCCATCGCCGAGGCCTGCCAGGCGTCGCCGC
GCGGC 225
A Y Y R T A A V K A A I E L G L F D V V G Q Q G R T P A A I A E A C Q A S P R
G 60

ClaI
ATTCGCATCCTTTGCTATTACCTAGTATCGATCGGTTTTCTACGCCGCAACGGTGGCCTGTTCTACATAGATCGCAACATGGCCATGTACCTGGATCGTAGTTCGCCCGGCTACC
TGGGT 345
I R I L C Y Y L V S I G F L R R N G G L F Y I D R N M A M Y L D R S S P G Y L
G 100

GGCAGCATCAAGTTCCTGCTCTCGCCCTACATCATGAGCGCCTTCACCGATCTGACCGCCGTAGTCAGGACCGGCAAGATCAACCTGGCGCAGGACGGCGTGGTGGCACCGGATC
ACCCG 465
G S I K F L L S P Y I M S A F T D L T A V V R T G K I N L A Q D G V V A P D H
P 140

CAGTGGGTGGAATTTGCACGCGCGATGGCACCGATGATGGCGCTGCCCTCGGCGTTGATCGCCAATATGGTGTCGTTGCCCGCTGATCGGCCGATTCGTGTGCTGGACGTGGCAG
CCGGC 585
Q W V E F A R A M A P M M A L P S A L I A N M V S L P A D R P I R V L D V A
A G 180
Motif
I
CACGGCCTGTTCGGCATCGCCTTCGCGCAGCGCTTCCGCCAGGCTGAAGTGAGCTTCCTGGACTGGGACAACGTGCTAGACGTAGCACGCGAAAACGCCCAGGCGGCCAAAGTGG
CCGAG 705
H G L F G I A F A Q R F R Q A E V S F L D W D N V L D V A R E N A Q A A K V
A E 220
(IR) HincII
CGAGCGCGTTTCCTGCCCGGCAACGCATTCGACCTCGATTACGGCAGCGGCTACGACGTGATCTTGTTGACCAACTTCCTGCACCATTTCGATGAGGTCGATGGCGAGCGCATCT
TGGCT 825
R A R F L P G N A F D L D Y G S G Y D V I L L T N F L H H F D E V D G E R I
L A 260
EcoRI Motif II
AAGACGCGCGATGCGCTGAACGACGACGGCATGGTGATCACTTTCGAATTCATCGCCGACGAAGAGCGTTCCTCACCGCCGCTGGCCGCCACCTTCAGCATGATGATGCTGGGCA
CCACC 945
K T R D A L N D D G M V I T F E F I A D E E R S S P P L A A T F S M M M L G
T T 300
Motif III ClaI
CCGGCGGGCGAGTCCTACACCTATAGCGATCTGGAAAGGATGTTTCGGCATGCCGGCTTCGGCCACGTGGAACTAAAATCGATACCGCCGGCCTTGCTGAAAGTGGTGGTTTCCC
GCAAG 1065

CGTTC
P A G E S Y T Y S D L E R M F R H A G F G H V E L K S I P P A L L K V V V S R
K 340

AGGGCCCCATAATGATCGAATCGGCGACATCCCCTGTGGCGAAAACCGAGCGCATCTGGTGCACCGAGCTGGACCTGGATGCACTCAACGCCATGTCGGCCAACACGATGCAGGC
CCTGC 1185
TCCCGGGGTATTACT (A3R)
R A P *
343

TCGGTATACGCATGATCGAGATCGGCTCGGACTATCTGGTCTCCTGCATGTCGGTGGACTGGCGTTGCCACCAGCCCTATGGGGTATTGCATGGCGGCGCATCGGTCACCCTGGC
CGAGG 1305

NotI
CTACCGGCAGCATGGCGGCCTCCATGTGCGTGCCGGCCGGCCAACGTTGCGTTGGCCTAGACATCAATGCCAACCACATCGCGAGCATCTCCAGTGGCCAAGTACAGTGCATCGC
GCGGC 1425
NcoI
CGCTGCACATAGGGGCCTTGACCCAGGTATGGCAGATGCGCATCTATGACGAAGGTGACCGCACGATCTGCGTGTCGCGCCTGACCATGG
1515

FIGURE 13

Xal-XabC	174	VLDVAAGHG	236	SGYDVILL	267	ALNDDGMVIT
Sgl-TcmO	173	FVDLGGARG	234	PRADVFIV	263	ALTPGGAVLV
Sgl-TcmN	331	IADLGGGDG	393	TGYDAYLF	423	IGDDDARLLI
Smy-MdmC	64	VLEIGTFTG	135	GAFDIVFV	159	LVRPGGLVAI
Mxa-SafC	63	TLEVGVFTG	134	GTFDLAFI	158	LVRPGGLIIL
Ser-EryG	85	VLDVGFGLG	149	ETFDRVTS	178	VLKPGGVLAI
Spe-DauK	183	VLDVGGGKG	254	RKADAIIL	273	ALEPGGRILI
Sal-DmpM	208	VVDIGGADG	269	GGGDLYVL	298	AMPAHARLLV
Shy-RapM	106	VLEVGCGMG	155	VQGDAEEL	194	ALRRGGALSH
Sav-AveD	71	VLDVGCGSG	124	GSFDAAWA	151	VLRPGGRLAV
		Motif I		Motif II		Motif III

FIGURE 14

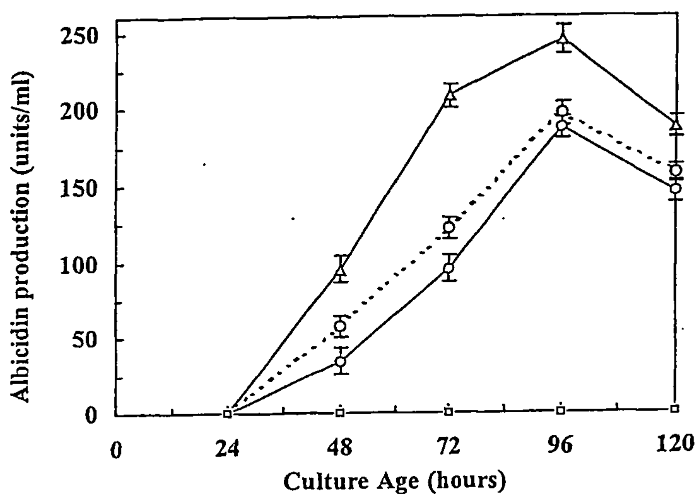

250
200
150
100
50
0
Albicidin production (units/ml)
0
24
48
72
96
120
Culture Age (hours)

FIGURE 15

INTERNATIONAL SEARCH REPORT	International application No. PCT/AU01/01190

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: C07K 14/195; C07H 21/04; C12N 15/52, 15/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN: File Reg, File CA (subsequence search of the individual sequences of Claim 1 combined with keywords ketoacyl reductase, polyketide, antibiotic, xanthomonas, albicidin in File CA), Index (CA, WPI, Medline, keyword xanthomonas albilineans, polyketide, gene)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STN FILE MEDLINE ABSTRACT 2001060291 & G. HUANG *et al.*, Gene, 255(2), September 19 2000, pp. 327-333. See abstract and GENBANK sequences AF239749 and AF238750 and CAS Registry number 332004-68-9.	1-94
X	PUBMED ABSTRACT 10780924 & F. SCHAUWECKER *et al.*, Chem. Biol., April 2000, 7(4), pp. 287-297. See abstract and GenPept sequence AAF42473, positions 947-952	1-94
P,X	STN FILE MEDLINE ABSTRACT 20011526256 & G. HUANG *et al.*, Microbiology, March 2001, 147(3), pp. 631-642. See abstract.	1-94

[X] Further documents are listed in the continuation of Box C [ ] See patent family annex

* Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 November 2001	18 DEC 2001
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer **L.F. MCCAFFERY** Telephone No : (02) 6283 2573

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/01190

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	STN FILE MEDLINE ABSTRACT 2001087312 & G. HUANG *et al.*, Gene, 258(1-2), November 27 2000, pp. 193-199. See abstract and CAS Registry number 346735-79-3.	1-94
P,X	STN FILE MEDLINE ABSTRACT 2001122130 & G. HUANG *et al.*, FEMS Microbiology Letters, December 1 2000, pp. 129-136. See abstract.	1-94
X	STN FILE CA ABSTRACT 132:103595 & K. MAYER et al., Nature, 1999, 402(6763), pp. 769-777. See abstract and CAS Registry number 254869-45-9	1-94
X	STN FILE CA ABSTRACT 130:120325 & S. T. COLE et al., Nature 1998, 396(6707), pp. 190-198. See abstract and CAS Registry numbers 208869-92-5 and 208869-94-7.	1-94
X	STN FILE CA ABSTRACT 123:331621 & F. BETSOU et al., Gene, 1995, 162(1), pp. 165-166. See abstract and CAS Registry number 170560-61-9.	1-94
X	STN FILE CA ABSTRACT 112:173135 & P. GLASER et al., Mol. Microbiol., 1988, 2(1), pp. 19-30. See abstract and CAS Registry number 126469-81-6.	1-94
X	STN 111:72007 & P. GLASER et al., EMBO J., 1988, 7(12), pp. 3997-4004. See abstract and CAS Registry 121889-91-6.	1-94

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/01190

Box I Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos :
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos : **1-94 (all in part)**
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
See attached sheet.

3. ☐ Claims Nos :
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

Supplemental Box
(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: I(2)

A subsequence search in the Registry file of the peptides defined by Claim 1 resulted in 2608 sequences that contained the defined sequences. This corresponded to 1381 Chemical Abstracts. Furthermore, this result does not take into account that the claims include variants and biologically active fragments of each of the individual peptides. Accordingly it is not economical to search this result.

Moreover, the invention appears to lie in the identification of new domains that are involved in Albicidin synthesis in X. albilineans. However, the domains defined in the claims constitute as few as four amino acids. Whilst this may be the active site of the domains in question, the function of a domain will be dependent on the constitution and topology of an entire region of the protein. Accordingly, the present search has been limited to the domains of the multifunctional polyketide-peptide synthase gene that appear to possess the functionalities defined in Claim 1, namely positions 1230-3116 (acyl-CoA ligase region); 3423-4724 and 9117-10367 (ketosynthase regions); 6660-7142 (ketoreductase region); 3117-3422, 8598-8795 and 8859-9068 (acyl carrier regions); 12447-14066 (adenylation region); 10890-11150 and 14067-14306 (peptidyl carrier regions); 11151-12446 and 14307-15632 (condensation regions). Similarly the search of the PPTase motifs have been limited to the specific peptides defines by SEQ ID NO 89, 93, 91, 87, 99, 101, 103, 105, 107 and 91. It is not possible and/or economically viable to search for variants and biologically acive fragments of these sequences.